

# The next big climate challenge

Governments should work together to build the supercomputers needed for future predictions that can capture the detail required to inform policy.

Few scientific creations have had greater impact on public opinion and policy than computer models of Earth's climate. These models, which unanimously show a rising tide of red as temperatures climb worldwide, have been key over the past decade in forging the scientific and political consensus that global warming is a grave danger.

Now that that consensus is all but universal, climate modellers are looking to take the next step, and to convert their creations from harbingers of doom to tools of practical policy. That means making their simulations good enough to guide hard decisions, from targets for carbon dioxide emissions on a global scale to the adaptations required to meet changing rainfall and extreme weather events on regional and local scales.

Today's modelling efforts, though, are not up to that job. They all agree on the general direction in which the climate will move as greenhouse gases build up, but they do not reliably capture all the nuances of today's climate, let alone tomorrow's. Moreover, each model differs from reality in different ways.

It was in recognition of this that a cross section of climate modellers gathered for a 'summit' at the European Centre for Medium-Range Weather Forecasts in Reading, UK, last week (see page 268). The meeting called for an ongoing project aimed at understanding and modelling the climate system well enough to provide the sorts of prediction that policy-makers and other stakeholders need — or, at the very least, to show why such prediction might not, in fact, be achievable. Key to this project would be one or more dedicated facilities offering world-class computational resources to the climate-modelling community.

## A clear resolution

Those resources are notably lacking at the moment. The world's very fastest computers run at hundreds of teraflops (which is to say, hundreds of trillions of mathematical operations a second), and the first forays into the petaflop range are expected by the end of the year. But today's climate models rarely run on machines that can manage more than a few tens of teraflops. This translates into spatial resolutions of a hundred kilometres or so. There was a general agreement at the summit that more realistic models will require resolutions in the tens of kilometres, at least. And even higher resolutions — a kilometre or less, say — may well be needed to handle such critical issues as cloud formation realistically. Hence the need for computers a couple of generations beyond the current state of the art.

Meeting this need is not just a matter of buying a supercomputer. It means moving climate modelling up the petaflop pecking order for a sustained period of time. One plausible goal might be to assure that the most powerful supercomputer in the public realm should be devoted to climate work by 2012, and that the field's lead should be regularly renewed for several computer generations after that. After

all, the fastest computers are nearly always paid for out of the world's public purses, often for use in areas of national security such as communications intelligence or nuclear weapons design. And climate prediction is a national security issue if ever there was one.

If funding agencies were to embrace such a goal, the implications would go well beyond money. Profound changes would be required of the community itself. Because the cost over a decade or more might easily top a billion dollars, such an investment in cutting-edge climate modelling would all but certainly have to be done multinationally, or even globally.

This would pull climate modelling into the world of 'big science' alongside space telescopes and particle accelerators — a transformation that would require new, and possibly disruptive, institutional arrangements.

## Living large

Aware of budgetary realities and the history of scientific centralizations, national centres of climate modelling and expertise such as Britain's Hadley Centre or the US National Center for Atmospheric Research might reasonably see the development as a threat. International collaborations such as CERN — the European particle-physics laboratory — and the European Southern Observatory have served the scientific communities of their member states well, yet have undoubtedly taken their toll on national facilities. With the upcoming inauguration of the Large Hadron Collider (LHC) at CERN, Europe will have the world's best particle-physics facility — but it will have very few of its other particle-physics facilities.

This analogy is not, however, fully convincing. Building an LHC does not make it significantly easier to build lesser accelerators. But advances in supercomputing do make it easier to build computers formerly known as super: a petaflop will seem slow in less than a decade. A world facility where teams of researchers try out very high resolutions and new techniques, and where software engineers and programmers learn how to get the most out of bleeding-edge hardware, will require a network of more modest centres around the world from which to draw its problem-list and into which to feed its insights. A range of operational climate-prediction capabilities could help keep modelling close to stakeholders' needs, and lessen the all-eggs-in-one-basket group-think risk of a global facility.

An ambitious climate-modelling facility dedicated to solving problems beyond the capability of today's national programmes carries risks, but they are risks worth taking. The world's governments — and even, conceivably, its high-tech philanthropists — should listen to the modellers. Big science is often, and gloriously, justified on the basis of pure intellectual excitement. This field offers that — and a chance to improve the world as well. ■

**"Climate prediction is a national security issue if ever there was one."**



## Stuck in the mud

The Environmental Protection Agency must gather data on the toxicity of spreading sewage sludge.

Some 30 years ago, as the United States began to tighten its environmental regulations on residential and industrial wastewater, operators of sewage-treatment plants embraced what seemed an eminently sensible idea. They decided to take the rich organic sludge left over after clean water is extracted and sell it to farmers as fertilizer.

The practice proved popular, and has become increasingly common internationally. Today, some 60% of sludges, innocuously dubbed 'biosolids' by the US Environmental Protection Agency (EPA), are used as fertilizer in the United States.

The programme might well be as sensible as it seems. It is possible that the millions of tonnes of sludge being spread across the rural landscape contain no significant levels of toxic chemicals, heavy metals or disease-causing organisms. It may all be perfectly benign. The disturbing fact is that no one knows.

In what can only be called an institutional failure spanning more than three decades — and presidential administrations of both parties — there has been no systematic monitoring programme to test what is in the sludge. Nor has there been much analysis of the potential health effects among local residents — even though anecdotal evidence suggests ample cause for concern.

In fact, one of the studies used to refute potential dangers,

published in the *Journal of Environmental Quality* in 2003 by researchers at the University of Georgia in Athens, has been called into question (see page 262). Even the National Academy of Sciences seems to have been taken in. A 2002 report from the academy cited the then unpublished Georgia work as evidence that the EPA had investigated and dismissed claims that sewage sludge had killed cattle, but the study had not looked at the dairy farms in question. And although it may be technically true that there was no documented evidence of sludge applications causing human illness or death, the academy also cited work by an EPA whistleblower, David Lewis, suggesting at least an association between these factors. If anything, recent research underscores those findings.

The Georgia citation notwithstanding, the academy did outline a sound plan for moving forward. It recommended among other things that the EPA improve its risk-analysis techniques; survey the sludges for potential contaminants; begin tracking health complaints; and conduct some epidemiological analyses to determine whether these reports merit concern.

The EPA has completed none of those tasks. Six years later, the agency is only now trying to finish its evaluation of potential contaminants and has yet to establish a system for monitoring reports of health problems. Agency officials say that they are working on risk-analysis tools, but have yet to undertake any kind of epidemiological studies.

The EPA certainly has other competing priorities, and the fault here does not lie only with the current administration or any single researcher. Regardless, these safety questions deserve answers, and the EPA should be able to deliver them. It is time to get the data. ■

## Negative results

Retracted papers require a thorough explanation of what went wrong in the experiments.

At first glance it seems to be a shining example of the scientific method in action. Two papers published by biochemist Homme Hellinga and his students at Duke University Medical Center in Durham, North Carolina, claimed a breakthrough in rational enzyme design. Last year, another chemist found that Hellinga's enzymes didn't actually work, which led to the retraction of the two papers this February (see page 275). Then, this March, a third group published research showing that rational enzyme design really is possible. All has ended happily, it seems, with the field marching forward in triumph.

But examined more closely, the episode reveals some less than happy aspects of science as it is actually practised. For example, the problems with Hellinga's enzymes were identified by John Richard at the State University of New York in Buffalo, who hoped to use the proteins in his own work. In effect, Richard and his two co-workers wasted seven months and tens of thousands of dollars failing to reproduce the results from Hellinga's lab. Richard's subsequent efforts to correct the scientific record thus came at considerable cost, with no discernable benefit to his own career.

This is a perennial problem in science. Many researchers who come across non-reproducible work save themselves extra hassle and money by simply not pursuing it further. Meanwhile, those who refuse to let it go — like Richard — gain nothing.

The process has been even more difficult for Hellinga's former student, Mary Dwyer. Hellinga accused Dwyer of faking data in the now-retracted papers, although he apparently had no evidence of intentional wrongdoing on her part. A Duke inquiry later cleared Dwyer of any misconduct, but her mentor's accusations could be more damaging to her career than the retractions.

The situation highlights the vulnerability of students in the system of scientific mentorship. Indeed, Hellinga's decision to accuse Dwyer was questionable. As Dwyer's adviser, Hellinga was responsible for training her. If she made mistakes, they are ultimately his responsibility. Instead, by accusing her, he cut off any possibility of frank and open discussion.

In the end, despite Richard's work, a misconduct inquiry and two retractions, the scientific community still does not know what went wrong, and may never know. This is perhaps most damaging to Hellinga, as the scientific community is now unsure whether to trust his prior and subsequent work. He and Duke owe the world a more thorough explanation of his conduct, and of his scientific work in question. Until then, this episode will continue to be a cautionary tale about the weaknesses — not the strengths — of the scientific process. ■



# RESEARCH HIGHLIGHTS

## NEUROSCIENCE

### Hearing what and where

*Nature Neurosci.* **11**, 609–616 (2008)

Neuroscientists have long thought that the brain uses different regions to locate sounds and to analyse them, as is known to be true for vision. Stephen Lomber of the University of Western Ontario in London, Canada, now provides behavioural evidence that this is so.

Lomber trained cats to locate the source of a sound, and then to discriminate patterns of sound. When he chilled the cats' posterior auditory fields, they got worse at pinpointing a sound's source; when the cats' anterior auditory fields were chilled, they were worse on the pattern discrimination task. The test supports the idea that the brain processes the 'what' and 'where' of sound in parallel.

## PALAEOCLIMATE

### Methane didn't act alone

*Geology* **36**, 315–318 (2008)

Methane outbursts from seafloor deposits are unlikely to have been the sole cause of an extreme episode of global warming around the time of the Palaeocene–Eocene Thermal Maximum some 55 million years ago.

Karla Panchuk of Pennsylvania State University, University Park, and her colleagues configured an Earth-system model with early Eocene geography to assess the involvement of potential carbon sources with distinct isotopic signatures.

Methane could have produced an observed negative carbon-13 isotope bump, but couldn't have caused the calcium carbonate dissolution in the oceans estimated from sediment layers. The team suggests that a carbon pulse of at least 6,800 gigatonnes — three times what methane could produce — is needed to reconcile the observations.

## GEOSCIENCE

### The dust settles

*Global Biogeochem. Cycles* **22**, GB2006 (2008)

Iron-rich dust, blown from arid lands and dropped into the oceans, is an important

nutrient for phytoplankton. But according to Thibaut Wagener at the Oceanographic Laboratory of France's National Centre for Scientific Research in Villefranche-sur-Mer and his colleagues, estimates for dust flux are 10 times too high in the Southern Hemisphere in global climate-change models.

They collected and analysed dust (see left, a silica particle entangled in sea salt) on cruises in the southern Pacific and Indian Oceans. The authors argue that the amount of iron in upwelling plumes of water limits phytoplankton growth in these oceans.

## QUANTUM OPTICS

### Open the box

*Europhys. Lett.* **82**, 30002 (2008)

How can online gamblers be sure that the casino isn't cheating? They can't — but the quantum gambling machine devised by Yi-Sheng Zhang and colleagues at the University of Science and Technology of China in Hefei could put an end to that. In their system,

the casino places a particle in a quantum superposition, where it is in two 'boxes' at once until the gambler opens one, 'collapsing' the particle into either box with equal probability. The gambler bets against the casino about which box the particle will be in.

The casino then sends the gambler the unselected box for checking: the quantum rules mean that if the casino tampered with the probabilities to bias the result, the gambler has a chance of finding out.

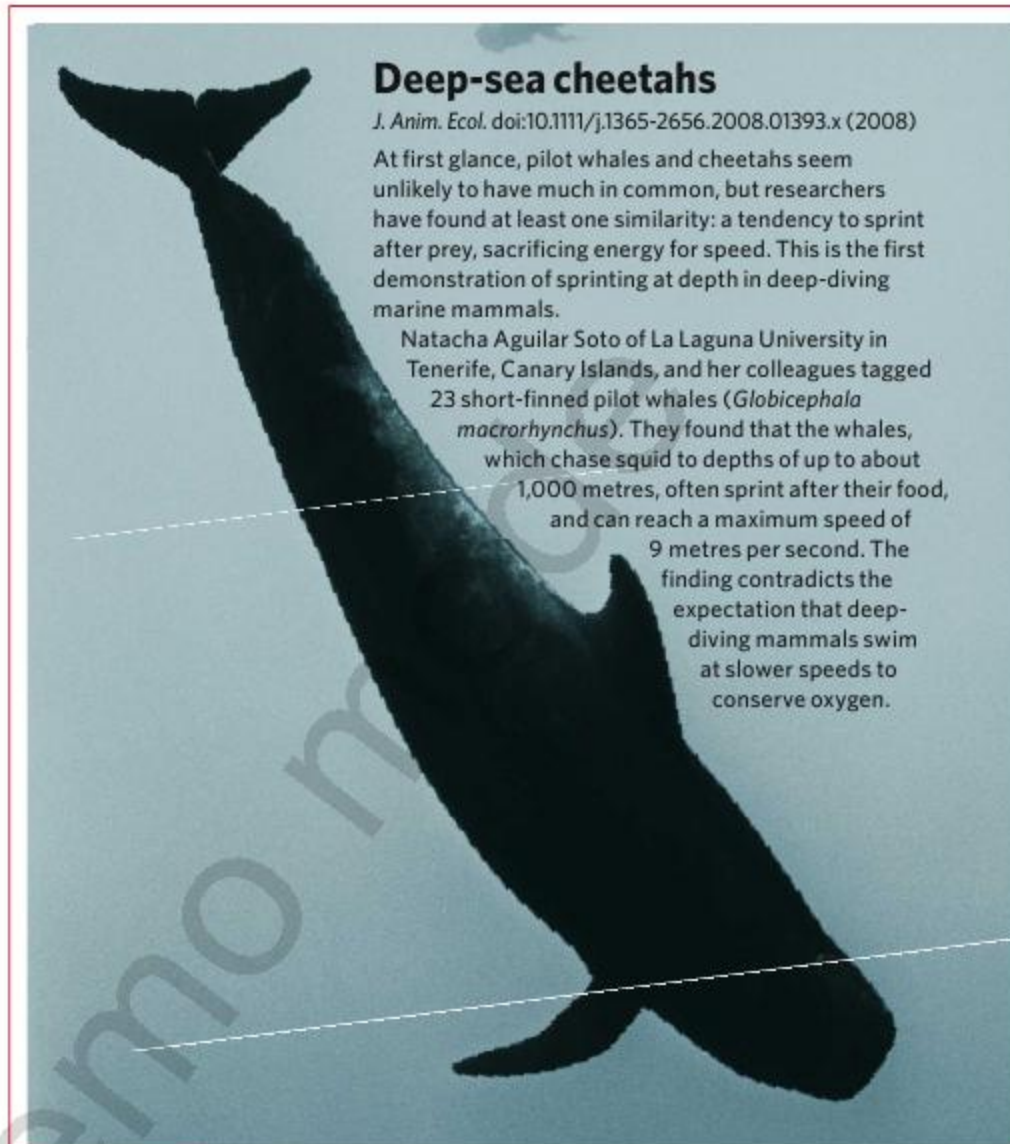
Zhang and colleagues have demonstrated the game using photons in superpositions of polarization states. Three detectors measure the result: two read the photon's state, and the third checks for tampering.

## NEUROBIOLOGY

### The heart in the head

*Science* **10.1126/science.1153651** (2008)

Do humans care more about distributing goods efficiently or fairly? Is this decision rational or emotional? Steven Quartz of

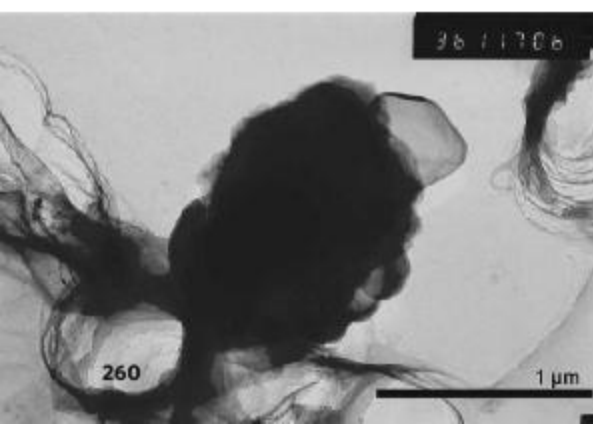


### Deep-sea cheetahs

*J. Anim. Ecol.* doi:10.1111/j.1365-2656.2008.01393.x (2008)

At first glance, pilot whales and cheetahs seem unlikely to have much in common, but researchers have found at least one similarity: a tendency to sprint after prey, sacrificing energy for speed. This is the first demonstration of sprinting at depth in deep-diving marine mammals.

Natacha Aguilar Soto of La Laguna University in Tenerife, Canary Islands, and her colleagues tagged 23 short-finned pilot whales (*Globicephala macrorhynchus*). They found that the whales, which chase squid to depths of up to about 1,000 metres, often sprint after their food, and can reach a maximum speed of 9 metres per second. The finding contradicts the expectation that deep-diving mammals swim at slower speeds to conserve oxygen.





the California Institute of Technology in Pasadena and his colleagues used functional magnetic resonance imaging to investigate.

The team made participants pretend to take meals away from orphans. They could take away either fewer meals in a more efficient distribution, or more meals in a fairer distribution. The researchers found that more efficient decisions correlated with more activity in a brain region called the putamen, whereas decisions that emphasized fairness correlated with more activity in the insula, which is involved with emotional processing. Differences in decisions came down to how averse participants were to inequity.

## ASTRONOMY

### A galaxy far, far away

*Asrophys. J.* **678**, 647–654 (2008)

Astronomers have spotted what seems to be the most distant galaxy ever observed.

The galaxy dates to 13 billion years ago, when the Universe was less than a billion years old. Larry Bradley of Johns Hopkins University in Baltimore, Maryland, and his colleagues spotted it by pointing the Hubble Space Telescope at a nearby galaxy cluster. The cluster's mass, about a thousand times that of the Milky Way, worked as a lens, magnifying the light from the galaxy behind it.

Bradley says the team now plans to try and obtain the galaxy's spectrum in order to confirm its distance. Such data could also provide important insights into how galaxies formed in the early Universe.

## ECOLOGY

### Hand-me-down bacteria

*Int. Soc. Microbial Ecol.* **2**, 510–518 (2008)

Developing earthworm embryos weed out unwanted microbes, while attracting their favoured bacterial symbionts to colonize their excretory organs.

*Eisenia foetida* bequeath their offspring a gift of *Acidovorax*-like bacteria, deposited directly into egg capsules where embryos will develop. Seana Davidson and David Stahl of the University of Washington in Seattle watched as the bacteria colonized the developing earthworm embryo. They found that the *Acidovorax*-like bacteria were selectively recruited to a small canal in each segment of the earthworm, where they lingered until excretory organs called nephridia developed sufficiently for the bacteria to colonize them. This selectivity for their symbionts suggests that the embryos fend off colonization by other microbes.

## ECOLOGY

### Hot chicks

*Science* **320**, 800–803 (2008)

A 47-year study of British great tits (*Parus major*; pictured below) shows that these birds can adapt to a changing climate purely through changes in behaviour, a phenomenon known as phenotypic plasticity.

Great tits at Wytham, near Oxford, now lay eggs an average of 14 days earlier than in 1961, keeping in step with the earlier profusion of moth larvae for feeding their chicks, report Ben Sheldon at the University of Oxford and his colleagues. This has been achieved purely through behavioural change rather than genetic selection, as shown by the fact that laying is tightly coupled to peak larval abundance even though springtime weather varies randomly from year to year.



## CHEMISTRY

### Disulphide dichotomies

*J. Am. Chem. Soc.* doi:10.1021/ja800180u (2008)

Chemists have measured how much a bond between two atoms stretches in the transition state, which is the short-lived phase between the initial and end states in a chemical reaction.

Sri Rama Koti Ainavarapu and Julio Fernandez at Columbia University in New York and their colleagues developed a force clamp technique to use in 'single-molecule force spectroscopy'. Using this, they held onto individual molecules and stretched them so as to trigger the reduction of a disulphide bond in a protein, invoking the transition state of the reaction. The researchers could then measure the distance between the sulphur atoms as they stretched apart.

## JOURNAL CLUB

Carl Bergstrom  
University of Washington,  
Seattle, Washington

**An evolutionary biologist considers the virulence of emerging infectious diseases.**

When a pathogen — for example, HIV — emerges into the human population, it adapts to growth and transmission in human hosts. At the same time, its virulence (often measured by case mortality) typically changes as well. On the basis of theoretical arguments and examples such as the myxoma virus, conventional wisdom holds that if a disease is highly virulent at first, it will rapidly evolve reduced virulence so as to maximize transmissibility. The idea is that pathogens face a virulence-transmissibility trade-off: strains that kill or even incapacitate their hosts are unlikely to spread as broadly as those that keep their hosts alive and mobile.

One might think — and some have argued — that we can take comfort from such reasoning. By this logic, the 60% mortality rate seen in human cases of H5N1 avian influenza should rapidly attenuate were a human pandemic to occur. But in the inaugural issue of *Evolutionary Applications*, Bull and Ebert refute this thinking using a clear, simple mathematical model (J. J. Bull & D. Ebert *Evol. Appl.* **1**, 172–182; 2008).

As someone working on the dynamics of emerging infectious diseases, I find this paper fascinating and sobering in equal measure. The gist of its argument is that trade-off models may not apply well to emerging infectious diseases, precisely because they are still emerging. When a disease first enters a new host, it can be far from the optimum point on the virulence-transmissibility trade-off curve. Its early evolutionary trajectory may be contingent on mutation supply and thus very hard to predict: virulence might decline, but could also initially rise.

The implications are clear. We need to invest now in disease surveillance, public-health infrastructure and pandemic planning. We cannot count on evolution to do our work for us.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS

Farmer Andy McElmurray won his court case against the US Department of Agriculture over land poisoned by sludge fertilizer.



R. EHRHARDT/AP

## Raking through sludge exposes a stink

A former US Environmental Protection Agency (EPA) scientist is suing the agency's officials and researchers at the University of Georgia in Athens, alleging that they manufactured and published false data to support the use of potentially harmful sewage sludges as fertilizers. The sludges have been linked to health problems in humans and cattle — and even deaths.

The False Claims Act lawsuit brought by microbiologist David Lewis, who says he was forced out of the agency, alleges that EPA officials and University of Georgia researchers fraudulently orchestrated a grant and then fabricated data to ensure that the EPA's 'biosolids' programme would come out smelling pretty. If the charges stick, the scientists and EPA officials could be held personally liable and may be forced to pay back the original grant as well as some US\$4.6 million in subsequent grants, plus penalties.

"This is one of the few ways that you can hold people accountable for bad science and indeed for using false information to create that science," says attorney Ed Hallman of Decker, Hallman, Barber & Briggs in Atlanta, who filed the lawsuit on behalf of Lewis and two Georgia dairy farmers.

At the heart of the case is a study by agricultural engineer Julia Gaskin of the University of Georgia and her colleagues, which concluded that using sludge as a fertilizer "should not pose a risk to animal health". It was used in a 2002

report by the US National Academy of Sciences (NAS), which brushed aside allegations that livestock had been killed by the toxic biosolids. The report states, with explicit reference to the Gaskin study, that the EPA had investigated these cases and found "no substantiation" to the allegations. Gaskin and her colleagues published their study a year later in the *Journal of Environmental Quality*<sup>1</sup>.

The lawsuit alleges that the researchers concealed their own evidence that sewage sludge applications contaminated land and probably contributed to cattle deaths on two dairy farms in Georgia, according to recently unsealed court documents. They then conducted a new study on different land — using sewage-sludge data that were known to be "fudged", in the words of one federal judge — to show that the use of biosolids is safe, according to the lawsuit.

Gaskin would not talk about specifics but says she stands by her work. She also says that

the paper was never intended to study problems with biosolids on the dairy farms. "The purpose of this paper was not the focus that has been alleged," she says. "That was not part of this effort."

University officials and the EPA declined to comment on the lawsuit or discuss the biosolids programme.

The US biosolids programme, which dates back to the 1970s, relies on residential and industrial wastes routed through thousands

of water-treatment plants. Some 60% of the residual sludges from the process — several million dry tonnes annually — are now used as fertilizers rather than being buried or incinerated. But questions remain about the sludges' impact on human and animal health — the programme has been the subject of multiple lawsuits for more than a decade.

### Court ruling

In February, a district court in Augusta, Georgia, ruled in favour of the McElmurray family, which had sued the Department of Agriculture for farm subsidies on land they could not plant because of various contaminants from sludge, including cadmium, molybdenum, arsenic and thallium. Judge Anthony Alaimo described a "broad consensus" that data on the city of Augusta's sewage sludge toxicity and its application were "unreliable, incomplete, and in some cases, fudged".

These were the same records that were used in the Gaskin study to calculate application rates on the farms that they analysed, and documents suggest that the researchers knew there were problems with the data. In one draft of the study, University of Georgia soil scientist William Miller scrawled a note with a smiley face saying: "We should fess up here that we don't know exact rates of application or specific characteristics of sludges applied."

Miller did not respond to e-mails or phone calls from *Nature*. In a recent interview with Associated Press, however, he acknowledged

**"Data on sewage sludge were unreliable, incomplete, and in some cases, fudged."**





**HAVE YOUR SAY**  
Comment on any of our  
news stories, online.  
[www.nature.com/news](http://www.nature.com/news)

these doubts but maintained that the study “does not include fake data”.

“I’m at a total loss to look at anything in the Gaskin paper or its conclusions that are not based on fabricated data or the concealment of their own data,” says Lewis, who claims he was forced out of the EPA in retaliation for his research into the health impacts of sewage sludge.

In 2002, Lewis and his colleagues published a study in the journal *BMC Public Health* documenting reported health problems among more than 48 people who lived near fields where ‘Class B’ sludges — the most common and least sanitized — were applied<sup>2</sup>. Some 25% of those surveyed were infected by *Staphylococcus aureus*, which contributed to two people’s deaths. This research was cited in the 2002 NAS report as well, although the report stated that there was no “documented scientific evidence” to substantiate reports of human illnesses or death. The academy said that it was not charged with evaluating human health claims but went on to acknowledge a “persistent uncertainty” about health impacts.

The NAS report recommended that the EPA conduct a new survey of chemicals and pathogens in sewage sludge, begin systematically tracking health complaints, and conduct epidemiological studies to assess the impacts of biosolids. The EPA has yet to implement these recommendations, although officials say a new survey of toxic chemicals found in sludges is due out later this year.

Last year, a team led by epidemiologist Sadik Khuder of the University of Toledo in Ohio published similar findings to those of Lewis’s team. Their larger study found that the risk of various health problems correlated with the proximity to farms where Class B sludges had been applied<sup>3</sup>.

“We have no idea what’s going into the waste-stream,” says Murray McBride, director of Cornell Waste Management Institute in Ithaca, New York. He says that there are unknown risks from cleaner ‘Class A’ sludges as well, because the sterilization process doesn’t kill all the pathogens and doesn’t affect a host of other chemicals used in modern industry. McBride says that the scientific community and regulatory agencies have been slow to address these questions because of the huge economic and institutional investment in the biosolids programme. “There’s a vested interest now in keeping this land application going,” he says. ■

**Jeff Tollefson**

1. Gaskin, J. W., Brobst, R. B., Miller, W. P. & Tollner, E. W. *J. Environ. Qual.* **32**, 146–152 (2003).

2. Lewis, D. L., Gattie, D. K., Novak, M. E., Sanchez, S. & Pumphrey, C. *BMC Public Health* **2**, 11 (2002).

3. Khuder, S. et al. *Arch. Environ. Occup. Health* **62**, 5–11 (2007).

See Editorial, page 258.

## German universities bow to public pressure over GM crops

Scientists have decried the decision by two German universities to pull the plug on field trials of genetically modified (GM) crops, calling it a “disgraceful” interference with scientists’ freedom to research.

“I am not happy at all with this decision,” says Stefan Hormuth, president of the Justus Liebig University in Giessen, Hesse. “Unfortunately, we were no longer able to deal with the massive opposition from politicians and the general public. The university has a reputation in the region that we cannot risk losing.”

Last month, the university announced that it would stop its planned cultivation of insect-resistant GM maize in nearby Gross-Gerau after activists occupied the 1,500-square-metre field. Another local field trial of GM maize, in Rauischholzhausen, was also stopped because of massive protests from the public and local politicians. Both trials had been approved by the national consumer protection and food safety body (BVL) and were to be conducted on behalf of Germany’s authority for agriculture variety and seed affairs.

Earlier in April, the rector and external advisory board of Nürtingen-Geislingen University in Baden-Württemberg “urgently recommended” that a faculty member stop his field trials on insect-resistant and fungal-resistant GM maize. The experiments, which were also approved by the BVL, had been going on since 1996. “We have always been very critical of this kind of research,” says economist Werner Ziegler, the university’s rector. “Lately things got out of control. There were e-mail attacks, vandalism, intimidation and personal threats. People started calling us ‘Monsanto University.’”

The final straw, Ziegler says, was when the local population brought food and blankets to activists occupying the university’s Oberboihingen test site. Local media and supporters hailed the illegal action as a brave act of civil inconvenience.

The university’s experiments were led by Andreas Schier, who studies fungal toxins in maize. Although legally the

university could not have forced him to stop the field trials, he says he eventually gave in because the pressure on him had become too great. “Scientifically, there was no reason whatsoever to discontinue the experiments,” Schier says. “But scientific arguments don’t count in a climate of mass hysteria.”

Schier claims that Ziegler and members of the advisory board threatened to publicly distance themselves from him and his research if he were to continue. “I couldn’t stand the pressure any more,” he says.

The incidents reveal a new level of public hostility to plant genetic engineering in

Germany, says Heinz Saedler, a director at the Max Planck Institute for Plant Breeding Research in Cologne, which this year is not cultivating GM crops either. “It is a very sad thing that some universities here haven’t got the backbone to withstand illegal activism and public pressure,” he says. “I honestly don’t have much hope left for the future of academic

research on GM crops in Germany.” “If it is indeed true that universities in Germany hinder faculty members from doing field research on GM crops for fear of being vandalized by anti-GM activists, then this is disgraceful,” says Vivian Moses, a visiting professor of biotechnology at King’s College London.

Vandalism and the destruction of GM crops have been common in Germany and elsewhere in Europe since field trials began 20 years ago. As a result, academic research in the field is becoming scarcer. Germany hosts around a third of the European field trials this year, on an area of just 30 hectares. Europe’s GM crop-cultivation research is almost negligible compared with that in the United States, Brazil and Canada.

“Work in the field is no longer appreciated because there is a perception that commercially it doesn’t lead anywhere, at least in the short term,” says Moses. “We need to face up to reality: is the global food crisis upon us, and must we take action, or will Europe continue to act as an ostrich, doing its best to ignore modern agricultural technology?” ■

Quirin Schiermeier



Andreas Schier had to stop his field trials of GM maize.

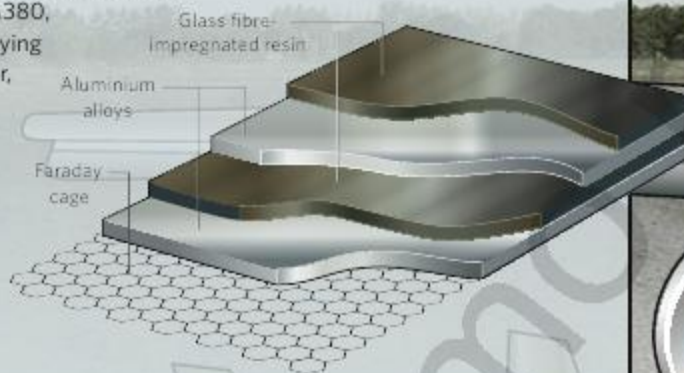


# Flights of green fancy

Air travel shows no sign of losing its allure but its environmental impact is not going to go away. **Katharine Sanderson** looks at some of the ways that scientists and engineers hope to reduce the carbon wing-print of aircraft.

Replacing all the metal in aircraft with lightweight composite materials could reduce fuel consumption by 25%, says aeronautical engineer Ian Poll of Cranfield University, UK. The fuselage of Boeing's 787 Dreamliner plane, to be unveiled in 2009, is made entirely of plastic reinforced with carbon fibre. Half of the aircraft's weight will come from composite materials. Airbus's megaplane, the A380, which has been flying since last October, has a composite material in sections of its upper fuselage. The material, called GLARE, is made from alternating thin layers of

aluminium and resin impregnated with fibreglass, making it more rigid and impact-resistant. GLARE could be combined with large sheets of conducting carbon nanotubes to protect against lightning. The nanotubes would be a lightweight replacement for the metal that lines the fuselage, and would form what's known as a Faraday cage.



## LOGISTICS

Streamlining air traffic globally could improve fuel efficiency by 12%, saving 73 million tonnes of carbon dioxide emissions a year, according to the Intergovernmental Panel on Climate Change. Geneva's International Air Transport Association claims that cutting flight times by 1 minute could save up to 4.8 million tonnes of CO<sub>2</sub> (and \$4 billion at current prices) a year.

## MATERIALS

## SIZE

Airbus's super-sized A380 can carry 853 people in economy class and has a fuel efficiency of 2.9 litres per passenger per 100 kilometres. The firm predicts that in the next 20 years, all classes of passenger planes will be 25% bigger than they are today. "We believe that one of the solutions to try to reduce the environmental impact of aviation is to have fewer flights," says Airbus spokesman Justin Dubon.

Conversely, global satellite positioning systems could lead to personal, autonomous airflight on

small planes, says Dennis Bushnell of NASA's Langley Research Center in Hampton, Virginia. Such planes could be powered by fuels that are less damaging to the environment than current sources.

The concept of a flying car remains popular, with many prototypes. Terrafugia, based in Woburn, Massachusetts, has a model called Transition, which has wings that fold out for flying and retract for driving (sketched above). Its car-plane should be available in 2009 for US\$148,000.

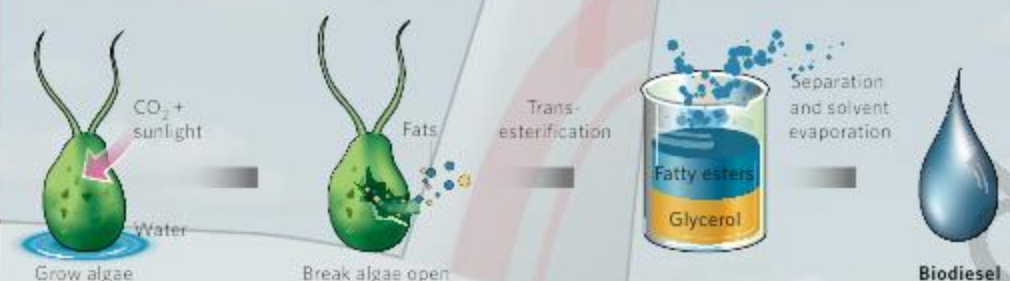




## FUELS

Algae look like the most likely candidate source for an aviation biofuel that won't compete with food for land. Algal biofuel can be made by cultivating high-fat-content algae to produce a biodiesel for use in a blend as jet fuel (see scheme). Other companies are looking at algal fermentation processes, hoping to make an aviation fuel directly.

In February, Airbus made a test flight with a synthetic liquid fuel made from natural gas. Such strategies pave the way to incorporating biomass-to-liquid fuels in the future, once they can be produced routinely from algae or other feedstocks such as jatropha or prairie grass.



Ross Walker, engineering programme manager for alternative fuel projects at Airbus, predicts that commercial flights will be powered by 25% biofuel by 2025. Whether 100% biofuel will be used is less certain "but it's an aspirational goal for the industry", says Boeing's Terrance Scott. UK entrepreneur Neil Laughton plans to fly his biofuelled, parasailing flying-car (pictured left) from London to Timbuktu, Mali, next year.

Nuclear power would remove the need for a gas turbine. The idea was first touted in the 1940s, and experiments were done by both the US and Russian governments during the cold war. Nuclear planes are dismissed by

some as too risky in case of crashes.

In December, APAME, an association promoting electrical aircraft based in St Pierre d'Argençon, France, flew Electra, the first conventional plane to fly on battery power alone, using lithium polymer batteries. Christian Vandamme, Electra's pilot, predicts that a battery-powered four-seater plane will be available in the next 10 years, and beyond that the development of a small commuter plane to carry up to 20 passengers.

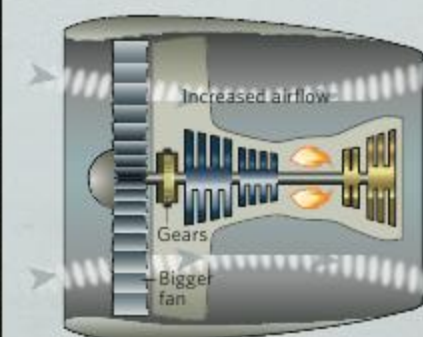
A plane powered by hydrogen — the European Union and Airbus Cryoplane project, launched in 2000 — never came to fruition because technologies to provide enough hydrogen for fuel are not yet available.

Boeing is testing fuel-cell technology using a small demonstrator aircraft, which is some 15–20 years away from public use. A fully solar-powered manned aircraft, *Solar Impulse*, plans to fly through the night, and round the world, in 2009 (see *Nature* 451, 884–886; 2008).

Others are investigating the possibility of using airships to transport freight around the globe.

## ENGINES

In today's engines, the turbine is a similar size to the engine fan, and they both spin at the same speed. But turbines are most efficient when spinning quickly, and engine fans are most efficient when going much slower. Introducing a series of gears allows the turbine to spin much faster than the engine fan, which can in turn be bigger without demanding much more power from



the turbine. A bigger fan ensures that most of the thrust comes from air that avoids the inner fuel-burning part of the engine — improving performance and reducing fuel burn. Engine manufacturer Pratt & Whitney in Connecticut is developing a geared turbo-fan that promises initial fuel savings of 15% and additional savings of 1% per year after the engines come into service. The engines have been snapped up by Japanese company Mitsubishi to use on the Mitsubishi Regional Jet, and Canadian aerospace company Bombardier will be using the engines on its CSeries. Both are due to enter service in 2013.

Another alternative is to use open-rotor, or unducted, fans with rotor blades that are free from any casing, removing a lot of drag from the system. But these fans would need to be large, potentially cancelling any fuel efficiencies by needing heavier propeller blades. They would also be very noisy.

## WINGS

The biggest chance to make aerodynamic improvements, and so save fuel, will come from new wing shapes. Blended wing body designs (see image below) are being developed to replace the tube-and-wing design now in use. A tubular fuselage is parasitic — it adds to the drag on the vehicle, but doesn't provide lift. Blended wing bodies remove the need for a fuselage: passengers can sit within the wing, as long as they don't mind not being able to see outside. The shape means that lift is generated over almost the entire structure. Drag is also reduced because the wetting area — the amount of the plane's surface in contact with the air — is reduced in a blended wing body compared with a conventional tube-and-wing design, so there is less friction between the wing and the air. The 'Silent' Aircraft Initiative by the University of Cambridge, UK, and the Massachusetts Institute of Technology in Cambridge, uses a blended wing shape to try to reduce noise to almost zero by improving aerodynamics and shielding the engines from the ground, which also saves fuel.



R. FERGUSON/BOEING/NASA





**A BROAD VIEW OF  
QUAKE 'PREDICTION'**  
Wide mapping of stress  
helps to pinpoint when and  
where the Earth will crack.  
[www.nature.com/news](http://www.nature.com/news)

PUNCHSTOCK

C. GUTIERREZ/UPH PHOTO/NEW SCOM

## SNAPSHOT

### Charged clouds

The Chaitén volcano in southern Chile erupts after more than 9,000 years of dormancy, spewing ash some 30 kilometres into the sky. Its plume stretches hundreds of kilometres across neighbouring Argentina to the Atlantic Ocean. The eruption has blanketed the surrounding province with a layer of soot, forcing the evacuation of thousands of people and threatening to bury the town of Chaitén, ten kilometres away. The spewed gases are high in silica and low in sulphur particles, so the eruption is unlikely to have a large effect on the climate.

For decades, scientists observing eruptions have seen lightning of the sort pictured here, but its origins are still not well understood. A study of an Alaskan volcano found that the ash particles spewing from the mouth of its crater were highly charged (R. J. Thomas *et al. Science* 315, 1097; 2007). Experts think that similar particles are behind these 'dirty thunderstorms'.

Geoff Brumfiel



## A side-splitting tale

Researchers who used a device to simulate human copulation report that more than 90% of condom breakages occur when the sheath is stretched repeatedly in the same place without returning to a relaxed state between stretches.

International standards for condom testing require two evaluations of the tensile strength of the condom material. One measures how far a ring of the material elongates when stretched between two rollers; the other involves inflating a condom until it bursts and recording the pressure and volume at which that happens.

But a sticky problem remains: clinical trials have reported failure rates of around 1% due to breakage, but the sample sizes were too small to reliably indicate why the devices break. Laboratory tests of condoms made of varying thicknesses and materials have done a poor job of duplicating real-world splitting, making it difficult to design safer condoms.

Now, a team led by Nicholas White, head of quality control at SSL International in Cambridge, UK, which owns major condom brand Durex, has attempted a more realistic model of sex with a condom, using a device with an adjustable 'thrust-hole' diameter, thrust rate and lubrication.

First the team used a microscope to examine the breaks in 972 male condoms made of latex or polyurethane that were returned to the company between 1998 and 2005 by dissatisfied customers. More than 60% of these condoms were broken at the closed end, often with an outwards circular rupture that the research calls an 'eruption', where there had been no manufacturing flaw in the condom film, nor any evidence of misuse.

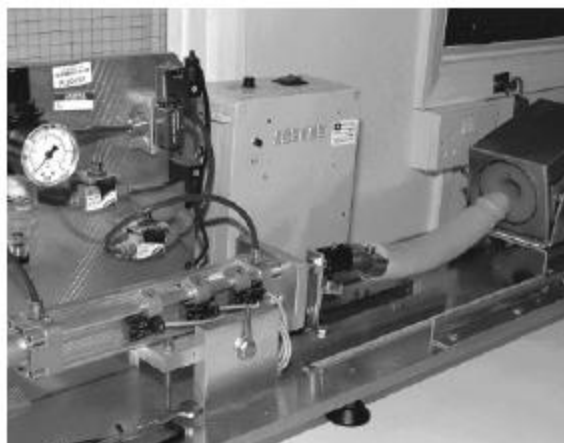
The researchers replicated this breakage pat-

tern by altering the parameters of their device so that the test condoms were progressively stretched at the tip during repeated thrusting. As expected, latex, from which the vast majority of condoms are made, showed a better elastic recovery than polyurethane, which is thinner and used in more expensive condom products. Yet both materials tended to break in the same way, irrespective of their elasticities (N. D. White, D. M. Hill and S. Bodemeier *Contraception* 77, 360–365; 2008).

"Just about the only thing that humans aren't strong enough to break, that is also thin and flexible enough, is parachute nylon — but that's porous," comments White. He hopes that the research will help in the design of condoms that can better withstand eruptions.

"Given the importance of condoms in stopping sexually transmitted infections as well as unwanted pregnancy, research to understand and then eliminate breakages is vital," says Lisa Power, head of policy at the Terrence Higgins Trust, an HIV charity in London, UK. "The safer we can make them, the happier a lot of people will be."

Anna Petherick



An apparatus simulating sex was used to test condoms.



# They say they want a revolution

Climate scientists call for major new modelling facility.

Climatologists have called for massive investment in computer and research resources to help revolutionize modelling capabilities. The eventual aim is to provide probabilistic climate predictions that are as useful, and usable, as weather forecasts.

At the end of a four-day summit held last week at the European Centre for Medium-Range Weather Forecasts in Reading, UK, the scientists made the case for a climate-prediction project on the scale of the Human Genome Project. A key component of this scheme, which would cost something up to, or over, a billion dollars, would be a world climate research facility with computer power far beyond that currently used in the field.

Questions on how severe the effects of global warming will be, and which regions will be hit in what ways, are beyond the capabilities of current climate science, at least in part because of computing constraints. Today's climate models are run on computers in the 10-teraflop range, meaning they are capable of 10 trillion operations a second. Despite this speed, models on these computers are still coarse-grained, cutting the world into cells more than 100 kilometres across.

Increasing computing power 10,000 times

— to speeds in the hundreds of petaflops — would allow modellers to study simulations at the kilometre scale, enabling better predictions on the activity of hurricanes and, eventually, the local deep convection that transfers much energy into the upper atmosphere (see 'A real solution?'). This research could then be fed into operational models.

The scientists think they could answer at least some of the 'big' questions on the effects of global warming if the technology was available. But national climate-modelling efforts, such as those of the Met Office in Exeter, UK, or the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, aren't attracting the required level of funding.

Although Japan's Earth Simulator in Yokohama was once the world's fastest computer, there are now 29 faster ones, with the first petaflop machines only months away.

"We need to be breathtakingly bold, frankly, in terms of some of the calculations that we're going to do in order to push the climate-prediction effort forward," says Leo Donner, a physical scientist at the Geophysical Fluid Dynamics Laboratory of Princeton University, New Jersey. Antonio Navarra, a climate modeller at the National Institute of Geophysics and

**"We need to be breathtakingly bold."**

— Leo Donner



Researchers from around the world gathered in Reading, UK, for the summit.

Volcanology in Bologna, Italy, spells out the implication: "We're reaching the point where national resources are insufficient to answer the scientific questions."

More money and cutting-edge challenges would also provide some hope of retaining highly trained programmers with expertise in climate modelling. Conference chair Jagadish Shukla of the Institute of Global Environment and Society in Calverton, Maryland, says this resource is "decreasing faster than the sea ice" as staff are lured from research by the financial rewards and job security provided by companies such as Google.

Addressing the summit on its opening day, economist Jeffrey Sachs, the director of the Earth Institute at Columbia University,

## A real solution?

Is the answer to climate prediction sitting in your pocket? Lenny Oliker, John Shalf and Michael Wehner of the Lawrence Berkeley National Laboratory in California think it could be. In a proposal discussed at the Reading climate-modelling summit (see main story) they suggest that the very small processors in mobile phones might be ideal components for very large climate computers — if 20 million of them could be wired together in the right way.

To run at the sort of kilometre-scale resolution that could accurately model cloud processes, they argue, a computer has to be able to run

at a sustained speed of around 10 petaflops, and a peak speed of perhaps 20 times that or more. If built with traditional high-performance chips such as AMD's Opteron or Intel's Xeon, such a machine would be extremely expensive and power-hungry — perhaps requiring as much as 100 megawatts. Processors developed for cell phones are small — less than a square millimetre in area — and frugal in their power requirements, needing less than a tenth of a watt each. These advantages, the researchers argue, far outweigh the slower speed at which such processors work and would permit

construction of a multi-petaflop computer that was much cheaper both to build and to run.

In some ways this is an extrapolation of the approach that IBM has taken to its successful Blue Gene line of supercomputers, which also rely on many relatively small and slow processors. But it goes further in the sheer number of processors and in an architecture designed specifically for the demands of climate calculations, rather than general-purpose computing.

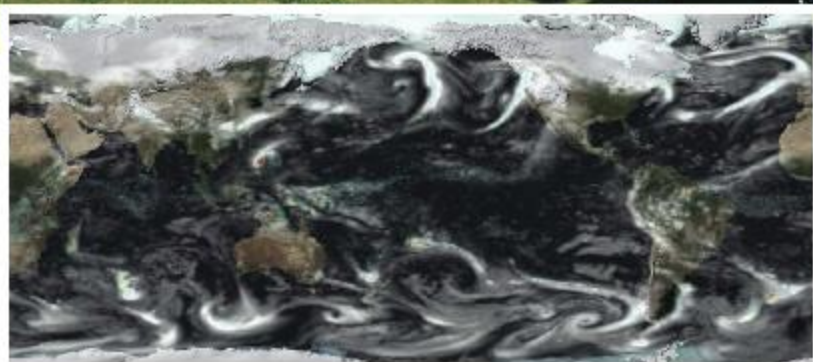
Per Nyberg of Seattle-based supercomputer makers Cray — which, like rivals IBM and NEC, sent speakers to the summit with an eye to business opportunities

— was understandably sceptical. "You can come up with back-of-an-envelope calculations about how cheaply you can build a computer but you have to be very, very careful." He argues that radical approaches can founder on software and on efficiency of usage, with a useful rate of number crunching far below the peak speed. And if the machine doesn't work as advertised, the expense of developing programs for it could be wasted. Wehner acknowledges the risks, but thinks careful prototyping and code development could minimize them. "We should build it and see," he says. **Oliver Morton**





R. HINE/ECMWF



State of the art: a model (left) from the UK National Centre for Atmospheric Science and the Met Office running on Japan's Earth Simulator.

P. L. VIDALE, NCAR/CLIMATE, WALKER INST., UNIV. READING

New York, said that there would be "a lot of interest among politicians in investing the hundreds of millions of dollars necessary, if scientists can provide answers to key questions... such as future food supply". Although governments are the obvious source of funding, Lawrence Gates, a now-retired climate scientist from the Lawrence Livermore National Laboratory in California, urged the attendees to explore philanthropic options.

How increased investment is divided between new facilities and existing ones is likely to be controversial. Some fear that a single global institute could threaten national centres, potentially taking the onus off governments to fund the institutions that are closest to stakeholders and could be expected to provide the predictions that have most real-world use. "Everyone is agreed that there needs to be a substantial investment in climate modelling, but whether a single centre is the solution is another question. There may be other ways," says John Mitchell, the chief scientist at the Met Office. Donner says that a sketch, presented on the conference's last day, of how the global facility might fit into the research world "seems to relegate national centres to little more than distributors of data".

However, Shukla was adamant that "every science breakthrough leads to the formation of an institute to address the problem". "We are facing

a darwinian change in the way we are working and we shouldn't be afraid of that," says Navarra. The meeting could have come to grief on such differences, according to Julia Slingo, the director of the Centre for Global Atmospheric Modelling at the University of Reading, but in the end the level of consensus, she says, was "fantastic".

Various attendees expressed frustration at the fact that the new facility could not be funded purely on the basis of the world-class science it would do — and indeed the fact that it would produce great research might count against it, making it seem more like a "toy for the boys" than a policy-informing instrument.

"If we just ask for enhanced understanding, then we have very little chance of getting the necessary funding," warned Shukla. But as Mitch Moncrieff from NCAR put it "we need a quantum leap in research to provide better predictions, even if the politicians don't get that". And there was widespread agreement that they need to get it fast. "We need a revolution as it has got to be done extremely quickly," said Brian Hoskins, director of the Grantham Institute for Climate Change at Imperial College London, UK.

Olive Heffernan

See Editorial, page 257.

## ON THE RECORD

**"It is important because if you're driving down the highway and you saw a crash of a small spaceship and a car or a bus full of kids, you really wouldn't know what to do."**

Denver citizen Jeff Peckman explains his proposed ballot initiative to create an Extraterrestrial Affairs Commission. The city is obliged to discuss the measure.

## SHOWBIZ NEWS

**This one's for you**

Canadian singer/songwriter Neil Young now has a trapdoor spider named after him: the *Myrmekiaphila neilyoungi* (pictured). Long may you run... away from it.



CECU NEWS SERVICES/REUTERS

## SCORECARD



### Long-lived data

An engineer has extracted data from a partially melted hard drive that fell to Earth after the *Columbia* shuttle explosion in 2003.



### Short-lived art

Curators at the New York Museum of Modern Art have had to kill an artwork consisting of a tiny jacket made of living stem cells after it grew too big.

## 3 GOOD REASONS

**... to send astronauts to an asteroid**

1. With the right asteroid it's easier than going to the Moon.
2. No one has been to an asteroid before.
3. Everyone loves *Le Petit Prince*.

A NASA study outlining such a plan is due to be published soon in *Acta Astronautica*.



Sources: Rocky Mountain News, Art Newspaper, Yahoo News, thestar.com

COURTESY OF GALLIARD EDITIONS

SIDELINES



## Space telescope lands new career in bomb detection

NASA's Compton Gamma Ray Observatory plunged into the Pacific Ocean, in a planned de-orbit, in 2000. But its spare parts are now being used to help detect dirty bombs.

James Ryan, an astrophysicist at the University of New Hampshire in Durham, has recycled parts from one of the telescope's old instruments. He aims to detect  $\gamma$ -rays emitted by radioactive substances, such as plutonium, uranium and caesium. These elements could be used in bombs that combine conventional explosives with radioactive material.

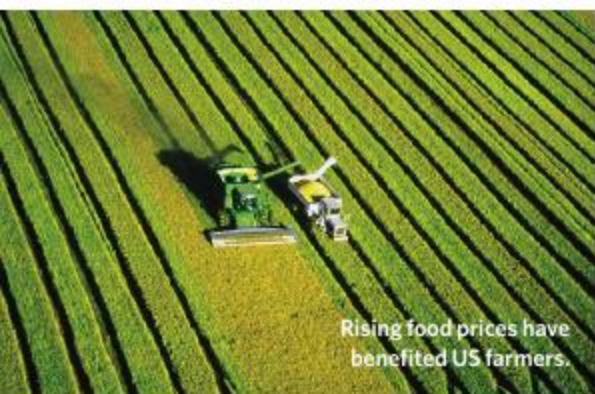
"If we can detect aluminium-26 on the other side of the Galaxy, we can detect this stuff on the other side of the street," says Ryan, who presented results from a prototype detector on 12 May in Boston, at a conference sponsored by the Institute of Electrical and Electronics Engineers and the US Department of Homeland Security.

## Farm bill reduces support for corn ethanol

US lawmakers have agreed on wide-ranging agricultural legislation that would reduce the federal subsidy on ethanol produced from maize (corn) by 12%. It marks the first move to scale back government support for a biofuel that has been linked to rising food prices and deforestation abroad.

The farm bill, worth nearly \$300 billion, would direct federal agricultural policy for the next five years, doling out money for agricultural subsidies, nutrition programmes, conservation projects and international food aid. It would reduce the maize ethanol subsidy from 51 to 45 cents per gallon; critics have called for its complete elimination. The bill also creates new incentives for the development of advanced biofuels that would not compete with food crops.

The bill is expected to clear Congress this week. But the White House says it will veto the bill because it fails to reduce agricultural subsidies at a time when rising crop prices are lining farmers' pockets.



Rising food prices have benefited US farmers.

## Forest clearance boosted power of Cyclone Nargis

As the death toll from Cyclone Nargis climbs to an estimated 100,000 or even higher in Myanmar, researchers are working to identify the factors that made it so deadly.

The category-4 Nargis, which struck Myanmar on 2 May, has far surpassed the death toll of Cyclone Sidr, which killed at least 3,500 people in Bangladesh last November, and is approaching that of the 1991 Bangladesh cyclone, which killed some 138,000.

Mangroves and coastal forests in Myanmar's Irrawaddy Delta have been extensively cleared in the past decade, mainly to make way for shrimp farms and rice paddies. The loss of that natural protective barrier intensified the destruction, conservation experts say. Areas where mangroves had been conserved saw less destruction than other regions in the 2004 Indian Ocean tsunami.



KHIN MAUNG WIN/AFP/GETTY IMAGES

## Cosmologist quits Britain over poor physics funding

A leading cosmologist is leaving the United Kingdom for Canada, in part, he says, because of the government's attitude towards funding physics research.

Neil Turok, a cosmologist at the University of Cambridge, says that recent funding cuts for fundamental physics in Britain "played a big role" in his decision to take up the post of executive director at the independent, non-profit Perimeter Institute for Theoretical Physics in Waterloo, Ontario. Turok says the cuts, and a desire for more applied research, are "the latest in a long history of the [government's] misunderstanding of the role of basic science".

Turok will bring valuable research and leadership experience to the institute, says Perimeter's founder, Mike Lazaridis.

## Hefty funds lay foundations for stem-cell facilities

The California Institute for Regenerative Medicine (CIRM), the state's stem-cell agency, has awarded \$271 million to 12 institutions to seed a building boom of research facilities across the state.

The awards were made at the CIRM's 7 May meeting. The largest, \$44 million, went to Stanford University in Palo Alto; the smallest, \$3 million, went to the University of California, Santa Barbara. Every University of California campus except Riverside received an award, as did the Buck Institute for Age Research in Novato; the University of Southern California in Los Angeles; and a consortium of four San Diego institutions called the San Diego Consortium for Regenerative Medicine (see *Nature* 453, 18–21; 2008).

California institutions hope the new

facilities, scheduled to be completed in two years, will help them recruit more scientific talent. The new buildings will be a boon for scientists working on human embryonic stem cells, but will also aid those studying other aspects of regenerative medicine, such as adult stem cells and immunology.

## Former NASA science director returns to post

Long-time NASA employee Ed Weiler has been named permanent director of the agency's US\$5-billion science division, returning to a position he held between 1998 and 2004.

Weiler, an astrophysicist, replaces Alan Stern, a planetary scientist who resigned in March after a dust-up with NASA chief Michael Griffin over the Mars exploration budget. Weiler himself is no stranger to such arguments; in 2004, he and Sean O'Keefe, NASA's then head, had differing views on whether a robotic servicing mission to the Hubble Space Telescope was a good idea. Later that year, Weiler took over directorship of NASA's Goddard Space Flight Center in Greenbelt, Maryland.

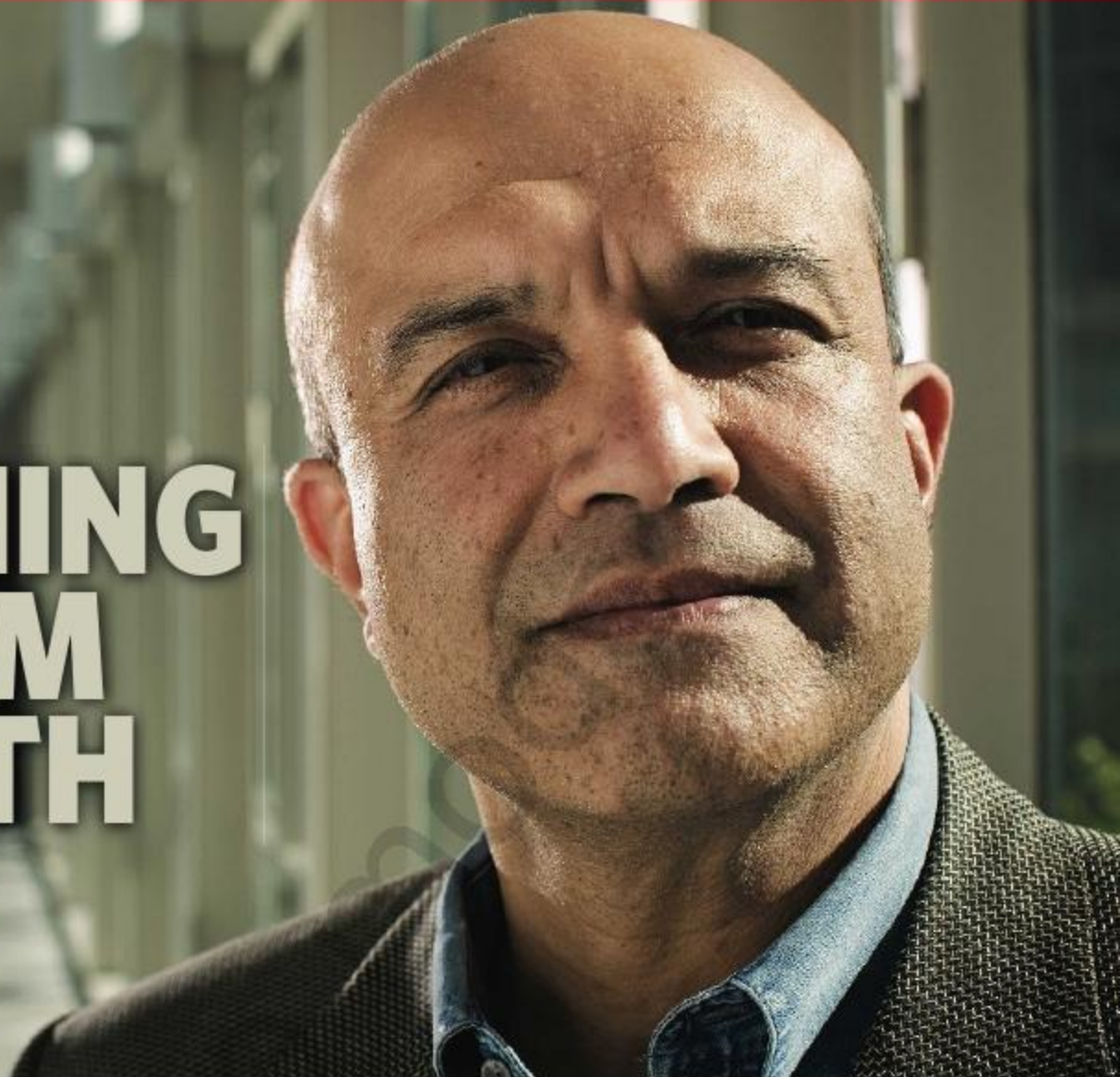
In general, Stern was praised by space scientists for his desire to boost launch rates and research funding. He also enforced a strict accountability for mission cost overruns (see *Nature* 448, 978; 2007), which ruffled some feathers, especially within the Mars community.

### NATURE.COM WINS A WEBBY

Nature.com has been named as the Best Science Website of 2008 in the 12th Annual Webby Awards ([www.webbyawards.com](http://www.webbyawards.com)). Hailed as the 'Oscars of the Internet' by *The New York Times*, the Webby Award winners are chosen by the International Academy of Digital Arts and Sciences (IADAS), a global organization of industry experts and technology innovators.



# LEARNING FROM DEATH



Vishva Dixit's study of cellular demise led to the discovery of a new molecular-signalling mechanism — one with implications for inflammation and perhaps much more, reports **Melinda Wenner**.

In a small, drab laboratory in Ann Arbor, Michigan, Vishva Dixit was getting bored. It was 1994 and although the University of Michigan associate professor had a grant to continue his studies on thrombospondin, a protein in the extracellular matrix, the work had become tedious. "I seemed to be dotting the 'i's and crossing the 't's," he recalls. "I felt that I needed to be doing something more important."

But what? Dixit had no idea. Biology's 'hot' topic at the time was the cell cycle — specifically, how cells grow and divide — but that didn't appeal to him. Then when reading the latest issue of *Scientific American* at home, he came across an article about cancer immunotherapy that briefly mentioned the role that tumour necrosis factors play in cell death. That instantly caught his attention. "I was always greatly intrigued by death," he says, but quickly recognizing how morbid that sounds, he adds, "Remember, I was trained as a pathologist."

Although he may overstate the role a magazine article played in his decision to change fields, it was a risk nonetheless. No one else at Michigan studied cell death at the time, and his own grant was intended to fund other

work. "If it didn't work out, I was done," he says — he wouldn't be able to renew the grant after having nothing to show for it. But that pressure, he says, galvanized him into action, and for the first time in years, he was excited. "I needed to be addressing a question that was important, where I could make some headway, where I felt I was having some impact," he recalls. "This was it."

## Big mysteries

Dixit's gambit paid off: his lab became famous for elucidating, in a series of landmark papers, what happens during cell death. Although he has now left academia — he is currently vice-president of research at Genentech, a biotechnology company based in San Francisco, California, and also oversees its postdoctoral programme — Dixit is still drawn to the big mysteries, especially when the rest of the field is preoccupied with other things. "The scientific question one addresses is as important as the answer one gets," Dixit says, choosing his words carefully, much as he chooses his projects.

Dixit has turned his attention to innate

immunity, winning recognition for his role in elucidating the function of the 'inflammasome', a complex of proteins inside immune cells that initiates attacks against many pathogens. He suspects that the inflammasome plays an important role in a number of different diseases, including type II diabetes. He is also helping to develop cancer therapeutics based on his research on cell-death signalling, several of which are now in early phase II clinical trials.

**"The scientific question one addresses is as important as the answer one gets."**


— Vishva Dixit

"He has had many different interests," says Andrew Chan, senior vice-president of immunology and antibody engineering at Genentech. "He tackles the big questions, goes after them with phenomenal rigour, and he perseveres." Dixit owns 53 patents and has published 124 journal articles, some garnering more than 2,000 citations.

Dixit's ambition, disregard for convention and love for science have their roots in his childhood. He grew up in Kenya, where both his North Indian parents were practising physicians at a time when, he says, only US\$1 was allocated per person per year for health care.

C. PICKENS





Genentech in San Francisco, California, is developing new anticancer drugs.

C. PICKENS

His mother told him stories about how drastically African medicine changed when the first antibiotic became available — before then, she and her husband told their dying patients that their best treatment was mountain air. “Then one day you have streptomycin. For people like her, it must have been such a miraculous day,” Dixit says. “It made me think about wanting to do something in science that would influence the lives of people.”

His interest in death also started in childhood, thanks to the forensic-pathology books that his parents kept at home. “For whatever bizarre reason, those books fascinated me,” Dixit recalls. “They had all these totally revolting pictures of gun-shot wounds, of decapitations that would be revolting to most kids, but that sort of captured me.” He decided to become a pathologist both because of his preoccupation with death and for practical reasons: it provided him with the most options. “Pathology traversed all medical disciplines,” he says.

Dixit wasn’t convinced, however, that as a

physician he would be able to make a real difference. It would be better, he thought, “to in some way contribute to the progress of medicine” or, in other words, to go into research. After completing his MD, Dixit left Kenya to go to Washington University School of Medicine in St Louis to pursue his residency along with a laboratory fellowship. After that, he was off to Michigan.

### A great big onion

When Dixit started studying programmed cell death, all that anyone knew about the process, he says, was that cells committed suicide after trigger molecules bound to death receptors on the cell surface. In a series of clever experiments, Dixit’s lab identified each component of the cell-death pathway — he named one of the death proteins ‘Yama’ after the Hindu god of death, although it is now known as caspase-3 — and determined how they were all connected<sup>1</sup>. “It was, in a way, a bit like peeling an onion,” says Jack Dixon, a pharmacologist at the University of California, San Diego, who was also at Michigan at the time. “Every layer that they took off revealed something interesting.”

One of the things that Dixit uncovered was an entirely new form of cellular signalling. Until then, scientists predicted that receptors signalled either by serving as ion channels or by adding and removing phosphate groups to

and from proteins in a continuous game of tag. A big question in the field was how, exactly, the death pathway started, and Dixit’s work revealed that death receptors use adaptor molecules to activate a cascade, initiated by protease enzymes called caspases, that ultimately dismantle the cell. In the “spectacular dance of death” that follows, Dixit says, swatches of the cell membrane ‘bleb’, bulging and blistering as they break away from the cytoskeleton; cavities form inside the cell; and the cell’s nucleus condenses drastically.

The signalling pathway was also far simpler than expected: Dixit predicted that multiple steps separated the adaptor protein from the activation of the death proteases, but his lab discovered that the adaptor directly interacts with the protease via a ‘death effector domain’<sup>2</sup>. “We realized the solution was embarrassingly simple and was staring us in the face,” he recalls.

Dixit’s work soon attracted interest from Genentech, the company that had given him, free of charge, his first batch of tumour necrosis factor. “They sent me a veritable tonne of the stuff,” he says. He moved to the company in 1997 because he was impressed by its dedication to basic research; he was also attracted by the possibility of translating his work into drug development.

Genentech has developed death-receptor activating factors such as Apo2L/TRAIL, which are being tested in combination with drugs against non-Hodgkin’s lymphoma. But Dixit largely has turned away from death. The

## A warming trend for the inflammasome

Coming in from the cold means more to some than others. For those with familial cold autoinflammatory syndrome, brief exposure to cold temperatures, even air conditioning, can result in fevers, rashes and other nasty reactions. In 2001, it and another rare inflammatory disease were linked to mutations in the aptly named protein cryopyrin. The discovery was expected to lead to many insights into immunology and inflammatory dysregulation<sup>6</sup>.

Indeed, a year later Jürg Tschopp, a biochemist at the University of Lausanne in Switzerland, showed that proteins like cryopyrin could become part of a large complex of proteins that spurs inflammation<sup>5</sup>. His group called it the inflammasome and have since linked its activity to those

rare inherited disorders and more common inflammatory reactions.

The inflammasome forms when an adaptor molecule such as cryopyrin, also known as NALP3, recognizes a toxic signal within the cell. These signals can be components shed by pathogenic bacteria or even the urate crystals that cause gout<sup>7</sup>. The signal allows the protein to recruit other adaptors ultimately forming a scaffold that activates caspases, protease proteins involved in cell death and inflammation (see diagram, opposite).

Three inflammasomes — NALP1, NALP3 and IPAF, named after the adaptors they use — have been identified to date, “but there probably are many others”, says Richard Flavell, an immunobiologist at Yale University School

of Medicine in New Haven, Connecticut. Each type responds to different threats, acting as a dedicated alarm system: the NALP1 inflammasome responds to *Bacillus anthracis* (anthrax); IPAF responds to *Salmonella*, *Shigella*, *Legionella* and *Pseudomonas*. Recently, Tschopp showed that NALP3 is activated by asbestos and is responsible for asbestos-related lung inflammation and fibrosis<sup>8</sup>.

However, little is known about what exactly each inflammasome detects. It’s possible that inflammasomes do not detect bacterial or viral products directly, but rather detect the secondary messengers produced by them. “That’s where the black box is,” says Vishva Dixit (see main story).

By interfering with the downstream effects of

inflammasome activation — specifically, by mopping up the immune regulator interleukin-8 — pharmaceutical companies including Amgen in Thousand Oaks, California, and Regeneron in Tarrytown, New York, have developed injectable medicines that show promise as treatments for inflammatory diseases. In a pilot study reported last summer, one such medicine, anakinra, made by Amgen, lowered blood sugar levels and improved  $\beta$ -cell function in 35 patients with type II diabetes<sup>9</sup>. “It could be that the inflammasome detects problems — too much glucose, for instance — and that triggers interleukin-1 $\beta$ , which then triggers type II diabetes,” says Tschopp. Whatever the reason, the field is hotting up.

M. W.



key questions have been answered, he says. He hasn't strayed far, though. When cells commit suicide, they build a protein scaffold that brings a caspase into contact with the death receptor. There are other types of immune-related caspases present inside the cell, too; Dixit wanted to elucidate their function, because he assumed they are also important.

Dixit and others, including Jürg Tschopp, a biochemist at the University of Lausanne in Switzerland, showed that just as the death caspases are recruited into an activating scaffold, other immune caspases — in particular, caspase-1 — integrate into an activating scaffold named the inflammasome<sup>3</sup> in 2002. This caspase activates immune factors called cytokines, which initiate an immune attack. In 2004, Dixit's lab provided unequivocal genetic evidence for the identity of specific scaffold components known as adaptors that are responsible for caspase-1 activation, and they showed that inflammasomes have the ability to distinguish between pathogenic attacks — for instance, between different types of bacteria — through the use of different adaptors<sup>4</sup>.

### Amazing instincts

The inflammasome may have important implications for medicine, particularly in treating inflammatory diseases, and maybe even in type II diabetes (see 'A warming trend for the inflammasome'). In the meantime, Dixit continues to follow his instincts, which, colleagues say, are amazing. "He has a really gifted sense of where to go with his science," says Jim Wells,

a cellular and molecular pharmacologist at the University of California, San Francisco.

For example, while studying how immune pathways turn themselves off, Dixit's group uncovered a surprising twist on a common cellular regulatory process called ubiquitination. Many proteins can be tagged by the addition of chains of protein labels called ubiquitins that are, as the name suggests, found all over most cells. Most of the time ubiquitination marks proteins as rubbish to be degraded and recycled, but in 2004, while studying a protein called A20, Dixit's lab noticed what it called 'ubiquitin editing', a more complex form of ubiquitin signalling<sup>5</sup>. Responsible for turning off the pro-inflammatory NF- $\kappa$ B pathway, A20 can both ubiquitinate and deubiquitinate the same proteins, a duality that at first seemed like "the ultimate futile cycle", Dixit recalls.

But ubiquitins can signal different things depending on how they are added to a protein. When A20 interacts with an immune mediator called RIP, it first removes an activating ubiquitin chain that is attached by one amino-acid residue (lysine) and then marks it for destruction by building a ubiquitin chain using a different lysine linker. "These seemingly contradictory activities actually act in concert to attenuate cytokine signalling," he says.

The man who started his career in a tiny run-down lab at the University of Michigan

may now have a roomy, windowed office with all the resources he needs at his fingertips, but colleagues say that he's still the same guy he's always been — brilliant, discerning, even a little bit calculating — although he is a much snappier dresser now. Dixit insists that his new environment has not changed his perspective on science: he still refuses to tackle anything but the biggest questions he can find, prepared to choose new paths whenever necessary. In fact, he doesn't even feel as if he has stepped out of academia completely

— Genentech, he maintains, is as strongly rooted in basic science as anywhere else, and he has at least as much freedom now as he did at Michigan. "People say, 'What's it like having gone to industry?'

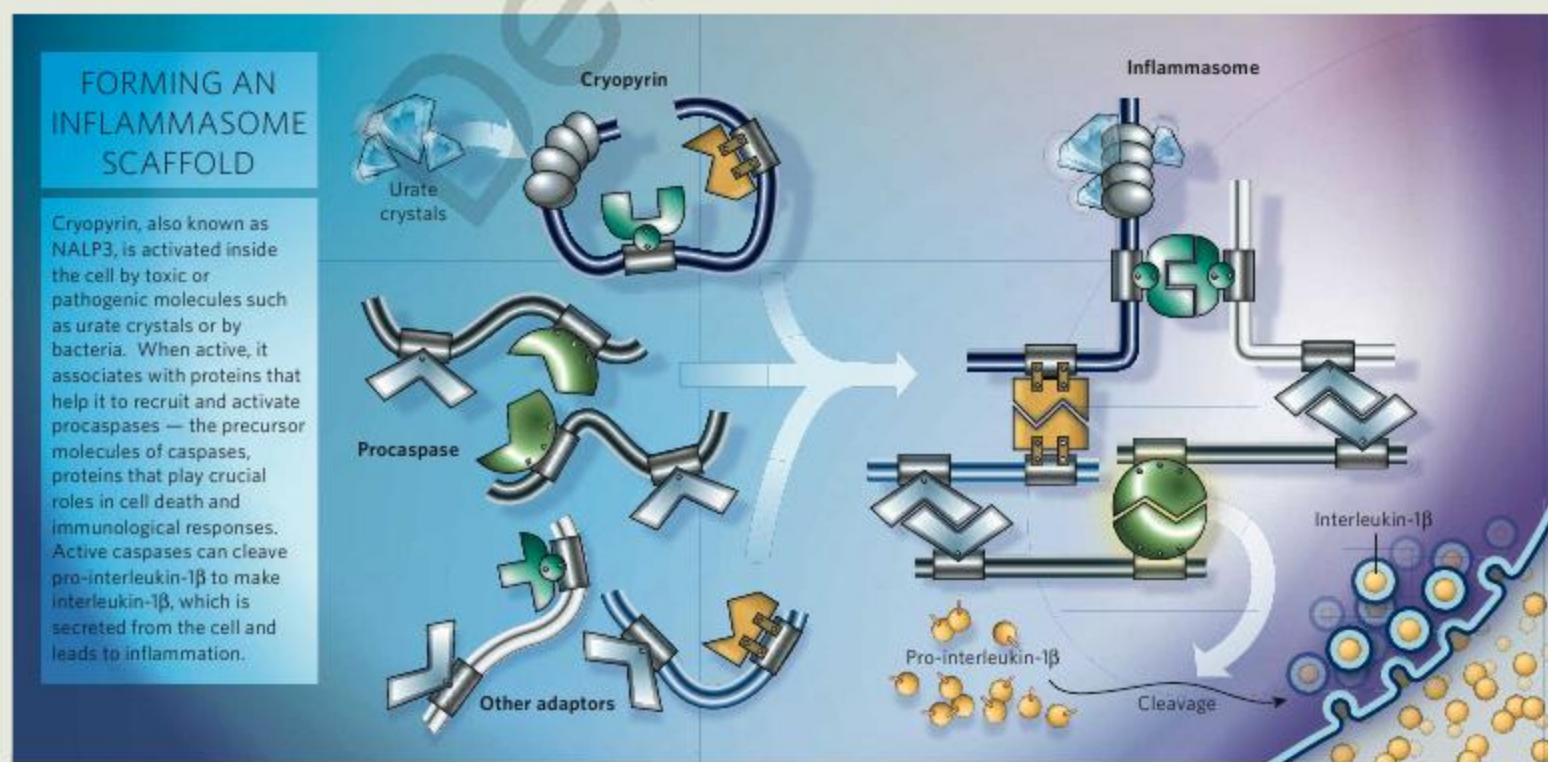
I say, 'To be honest, I really don't know, because I work at Genentech.'" One gets the feeling, however, that Dixit would feel at home anywhere, as long as he had the tools to address life's next big mystery.

**Melinda Wenner is a freelance writer in New York City.**

**"It was, in a way, a bit like peeling an onion. Every layer revealed something interesting."**

— Jack Dixon

1. Tewari, M. *et al.* *Cell* **81**, 801–809 (1995).
2. Muzio, M. *et al.* *Cell* **85**, 817–827 (1996).
3. Martinon, F., Burns, K. & Tschopp, J. *Mol. Cell* **10**, 417–426 (2002).
4. Mariathasan, S. *et al.* *Nature* **430**, 213–217 (2004).
5. Wertz, I. *et al.* *Nature* **430**, 694–699 (2004).
6. Hoffman, H. N. *et al.* *Nature Genet.* **29**, 301–305 (2001).
7. Martinon, F. *et al.* *Nature* **440**, 237–241 (2006).
8. Dostert, C. *et al.* *Science* **320**, 674–677 (2008).
9. Malozowski, S. *et al.* *N. Engl. J. Med.* **357**, 302–303 (2007).





# DESIGNER DEBACLE

A high-profile scientist, a graduate student and two major retractions. **Erika Check Hayden** reports on a case that has rocked the chemistry community.

When Mary Dwyer was looking for a doctoral adviser, Homme Hellinga was her first choice. A biochemist at Duke University Medical Center in Durham, North Carolina, Hellinga had ground-breaking ideas and an exciting research programme. He also shared Dwyer's interest in the relationship between protein structure and function. But there was a problem: students in Hellinga's lab were warning Dwyer away. "It's pretty tough," they told her; "there are other good labs." One student even pulled her aside and told her flat out that working with Hellinga was so difficult that she should not join the lab. By that time, that student remembers, many more students had left Hellinga's lab than had earned doctoral degrees under his tutelage.

Yet Dwyer had done a short rotation with Hellinga's group, and had seen nothing alarming. "I felt like I would probably be able to handle it," she recalls — and so, about nine years ago, she decided to join the lab.

Dwyer's work under Hellinga led to major publications in journals including *Nature* and *Science*, adding sparkle to Hellinga's already shining career. But last year, another scientist found problems that forced the eventual retraction of two papers — and Hellinga turned on Dwyer, accusing her of fabricating data. The episode has sparked controversy and condemnation, while highlighting the pressures on scientists working in cutting-edge research.

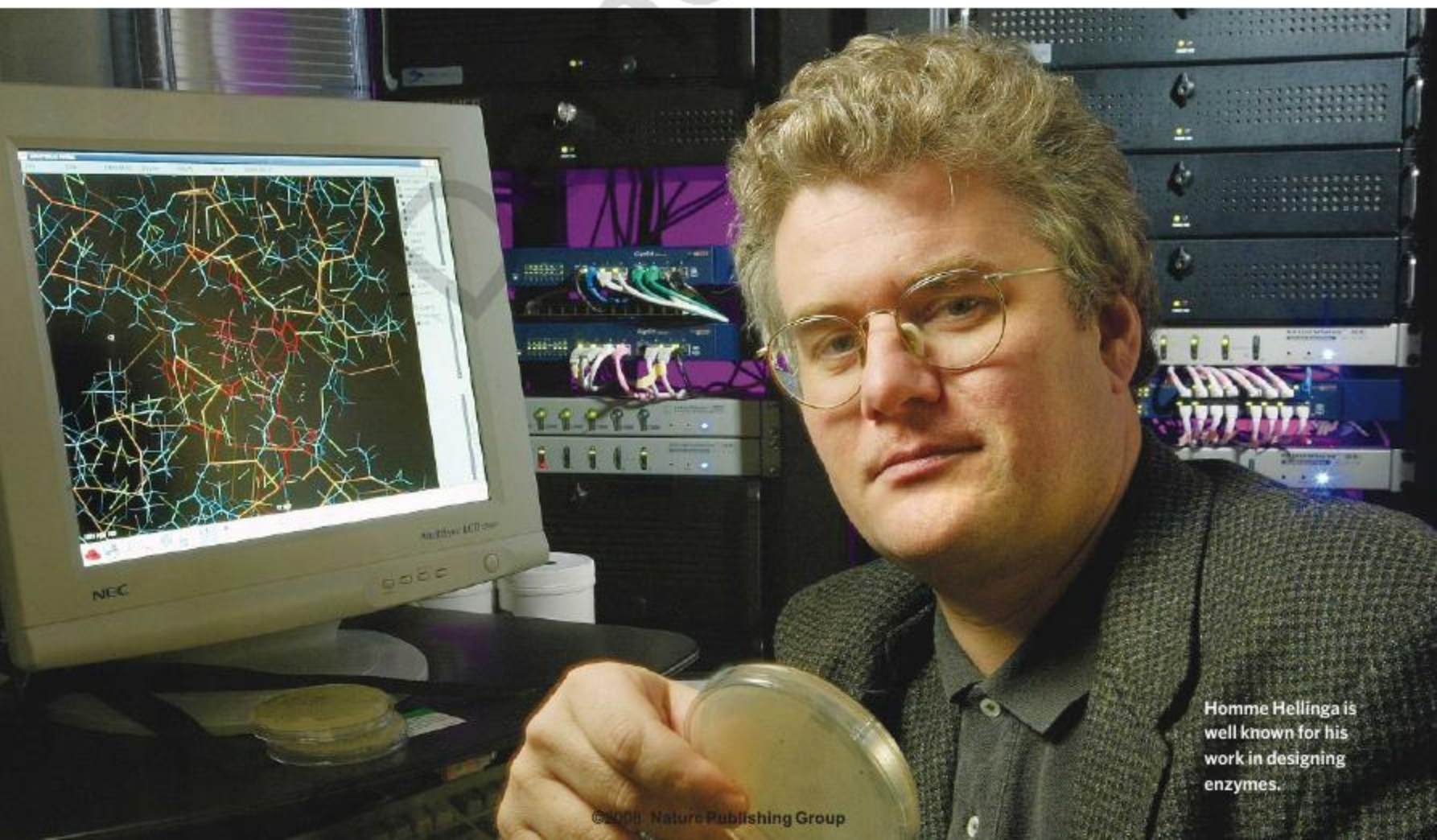
Hellinga is a bold scientist with a sterling pedigree. From his first *Nature* paper<sup>1</sup> onwards, Hellinga has been fascinated by one question: how does a series of amino acids encode a protein's function? Cracking that code is one of the major goals of science, because it would enable researchers to design custom proteins. In 1991, Hellinga, together with his postdoctoral mentor Frederic Richards of Yale University in New Haven, Connecticut, published a computer program<sup>2</sup> intended to do just that. Called DEZYMER, the program predicts protein sequences that

might adopt target structures and functions — some of which are new to nature.

It was fitting that Hellinga should take on such a problem. Those who know him describe Hellinga as highly confident in his intellect and interested only in grand challenges. One scientist recalls, for example, that Hellinga once asked a companion, "Do you think I'll be more famous than Darwin one day?" Asked whether he agrees with claims that he is arrogant, Hellinga replies, "I would say no. Can I appear to be personally arrogant? I would imagine yes. When you are trying to do a difficult experiment, you have to have a certain amount of self-confidence to say, 'All right, this is the moment and we think we have the techniques and ideas together to try and give this a go.'"

## Shapely targets

Around 2002, Hellinga decided to embark on his most difficult challenge yet: radically reshaping a humble protein into a highly active enzyme — a biological catalyst — called triose



Homme Hellinga is well known for his work in designing enzymes.



phosphate isomerase (TIM). The enzyme is part of a biological chain of reactions called the glycolysis pathway that is found in most organisms. Hellinga's goal was audacious; other scientists had designed weak enzymes<sup>3</sup>, but nothing as active as TIM — considered a 'perfect enzyme' because of its extremely high efficiency (see graphic).

Hellinga chose Dwyer and another student, Loren Looger, to work on the project in *Escherichia coli* bacteria. The pair were to transform *E. coli*'s ribose-binding protein, which has no enzymatic activity, into a TIM. Looger and Hellinga wrote computer programs to model how the structure of the ribose-binding protein could be changed to make it work like a TIM. Dwyer used the program to design mutated ribose-binding proteins, dubbed "NovoTIMs", and tested whether they worked in the lab.

Dwyer, who describes herself as a "pretty conservative person", was sceptical that the project would pan out. "I had my doubts all the time," she says. After about 6 months testing 25 designs, Dwyer found that a couple of the designed proteins were active, but she also noticed some problems. The *E. coli* bacteria made much smaller amounts of the NovoTIM proteins than of their own natural, or native, proteins. And the NovoTIMs were very unstable.

Perhaps because of these issues, Dwyer's experiments yielded confusing data about NovoTIM activity. When she measured the enzymes' kinetic parameters — characteristics that describe how enzymes work — the tests didn't always give the same results. "I felt like we couldn't nail down the kinetic parameters because of the variability that we were seeing," Dwyer recalls. Even after she started working with another member of the lab, "we were also getting a lot of variability. We just didn't understand it," Dwyer says. Hellinga says that the variability was "no more than you would expect in [such] an experiment".

By early 2004, Hellinga was ready to publish. On 29 March, he submitted a paper describing the NovoTIMs to *Science*, which accepted it on 6 May. The paper did not mention the variability Dwyer had noticed. It included only her best data and claimed victory<sup>4</sup>. "We have successfully converted a protein devoid of catalytic activity into a triose phosphate isomerase, using computational design techniques," it stated.

Dwyer was the first author on the *Science* paper, which was co-authored by Hellinga and Looger, who left Duke that year and now works at the Howard Hughes Medical Institute's Janelia Farm in Virginia. But Dwyer did not celebrate the accomplishment. "It was kind of strange,"

she recalls. "I wanted to work more on the variability issue," along with other odd results she had seen. "I felt like we weren't quite there yet."

Dwyer says that she raised her concerns with Hellinga at the time. But Hellinga says he does not feel he pushed Dwyer or anyone else to publish prematurely. "These things were talked through very carefully with all the people involved," he says.

That September, the National Institutes of Health gave Hellinga one of its nine inaugural Director's Pioneer Awards, worth US\$2.5 million over five years. In October, he received the \$10,000 Feynman Prize for experimental work from the Foresight Nanotech Institute in Palo Alto, California. Around the same time, he says, he and his wife, Duke structural biochemist Lorena Beese, were considering multiple job offers, including one from Yale. But in April 2005, Duke named Hellinga a James B. Duke Professor of Biochemistry, and Beese received the same honour the following year. Duke also created a new institute co-headed by the couple, the Institute for Biological Structure and Design.

### To the letter

As Hellinga's career was skyrocketing, it was perhaps easy for him to overlook a letter that crossed his desk in December 2004 amidst the flurry of accolades. "Dear Professor Hellinga," it began. "I was wondering if you would be interested in collaborating."

The letter was written by John Richard, a chemical biologist at the State University of

New York in Buffalo. Richard had studied with giants of the enzymology field: Perry Frey at the University of Wisconsin-Madison, Bill Jencks at Brandeis University in Waltham, Massachusetts, and Irwin Rose, now at the University of California, Irvine, who shared the 2004 Nobel Prize in Chemistry for discovering how a protein called ubiquitin marks other proteins for destruction in cells.

Richard had developed a method to analyse reactions catalysed by TIM<sup>5,6</sup>. He had seen Hellinga's *Science* paper and wanted to compare the characteristics of the NovoTIMs with those of normal TIMs. Richard

proposed such experiments to Hellinga, but received no response. "It wasn't a high priority," Hellinga says.

The two men come from very different scientific cultures. Richard was trained in mechanistic enzymology and is known for his work in physical organic chemistry — fields that are no longer in vogue, perhaps because "all the easy experiments have been done", Richard says. Richard has gained respect in these fields, which require carefulness and meticulousness. "John is clearly one of the best physical organic chemists in the world today working on enzymes," says Joseph Kappock, a biological chemist at Purdue University in West Lafayette, Indiana. By contrast, protein design — a hot field — requires daring, as it seeks not just to understand nature, but also to improve on it.

In July 2006, Richard was discussing the *Science* paper with another chemist, Jack Kirsch, an emeritus professor at the University of California, Berkeley, where Hellinga had given a seminar on his work. On 9 August, Kirsch sent Hellinga an e-mail. "[Richard] informed me recently that he had sent you an e-mail requesting materials," Kirsch wrote. "Is there any reason why you cannot comply with his request?"

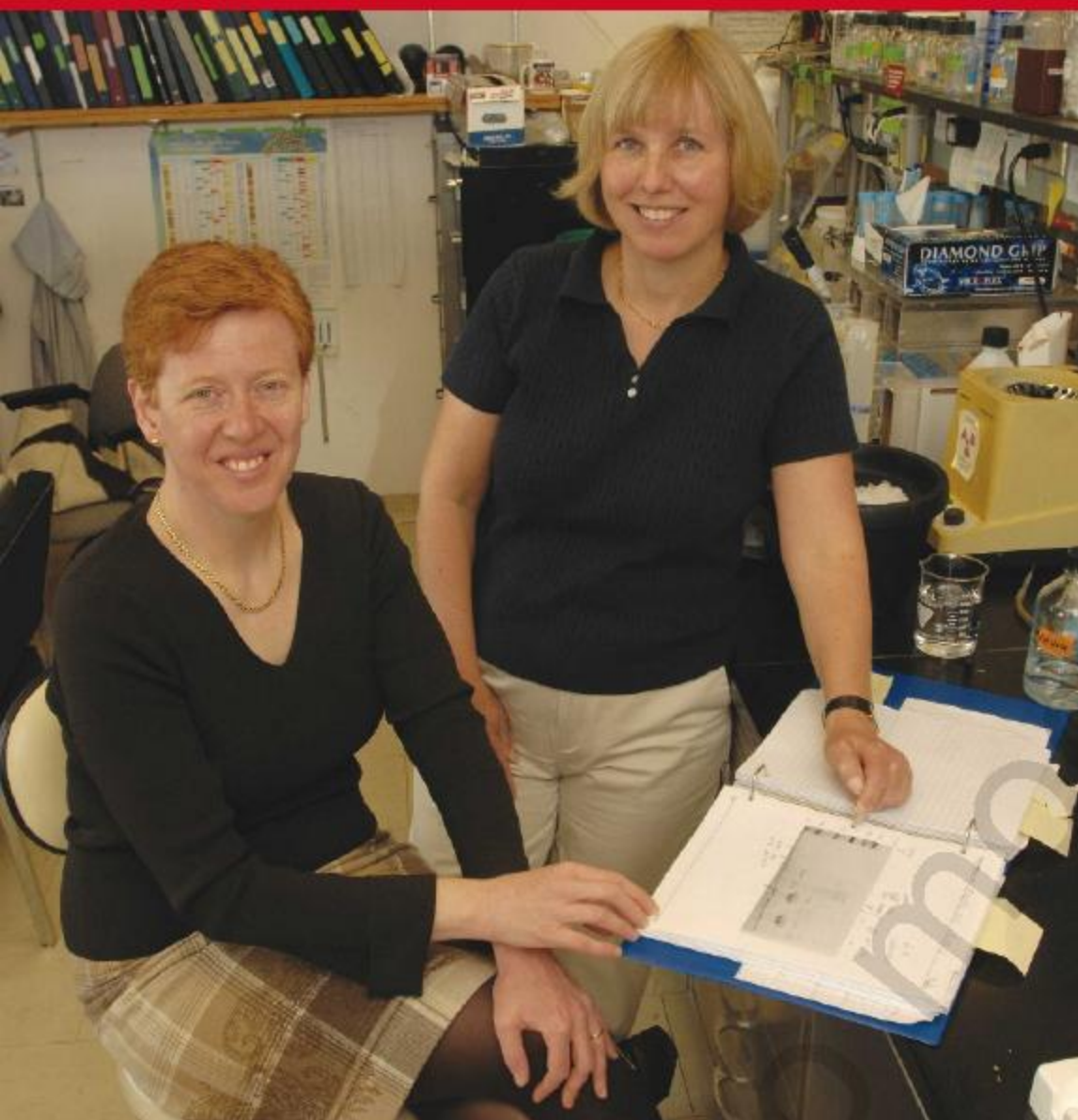
That e-mail seemed to grease the wheels. On 20 October, Hellinga wrote to Richard, agreeing to send DNA templates for the NovoTIMs he had made for the *Science* paper. He also sent templates for a second batch of NovoTIMs made by Dwyer and another researcher the year before. A paper describing these new proteins was about to be published in the *Journal of Molecular Biology*<sup>7</sup>. Hellinga sent Richard instructions for expressing and purifying all the NovoTIMs, as well as a note: "I hope that your experiments will be successful, and look forward to seeing the profiles for these designs."

In Buffalo, Richard hired a technician, Astrid Koudelka, to work on the NovoTIM project. Koudelka followed

**"It is a bush-league error not to purify your proteins well."**  
— Wallace Cleland







Tina Amyes (left) and Astrid Koudelka were unable to replicate Hellinga's work.

N. J. PARISI

Hellinga's notes, which instructed her to purify the NovoTIMs using a method called step gradient elution. But there was a problem: the step gradient could not separate the NovoTIMs from other contaminating proteins.

Then Richard's wife, chemist Tina Amyes, measured a kinetic parameter of the NovoTIMs — a value called the Michaelis constant. She found that it was different from the one reported in Hellinga's *Science* paper, but similar to that of natural, or wild-type, *E. coli* TIM. As Amyes studied the NovoTIMs throughout the first half of 2007, nothing about them was as Hellinga had reported, and everything suggested that they were wild-type TIMs.

Koudelka then modified Hellinga's procedures by using a continuous gradient elution, a more powerful purification method than the step elution. The new method cleanly separated the NovoTIMs from the contaminants. But when Amyes analysed the pure NovoTIMs, they had no enzymatic activity. Instead, the contaminating proteins were active — and looked just like wild-type *E. coli* TIMs.

By last July, the Buffalo group was convinced that something had gone wrong with Hellinga's experiments. By using step purification, they felt, Hellinga's lab had failed to separate the NovoTIMs from the TIMs found naturally in

*E. coli*. The NovoTIMs were inactive; instead, all the activity that Hellinga had reported in his papers was probably due to contaminating wild-type TIM. "I was sort of distressed," says Richard. "We spent quite a bit of time, money and resources to basically do nothing, to show something was wrong." Yet the team felt an obligation to try to correct the scientific record. "Just saying, 'This is not right, let's discard it and move on' — that's not fair to the scientific community," Koudelka says.

### Quick response

On 26 July, Richard sent a long e-mail to Hellinga that laid out his team's evidence, and pointed out what he saw as additional problems in some of Hellinga's other papers. Richard copied in the editors of the *Science* and *Journal of Molecular Biology* papers and two other chemists. "I think that these issues need to be dealt with in an expedient manner," Richard wrote, adding, "Please understand how difficult it has been for me to write this letter."

This time, Hellinga responded quickly. In a 30 July e-mail, Hellinga wrote that the key experiments "have been repeated several times by different individuals in my research group". The experiments included the tests that detected NovoTIM activity, and a set of negative control

experiments. These negative controls — not shown in either paper — found no activity in purified ribose-binding proteins, Hellinga said. But he agreed to look again at the NovoTIMs: "We will carry out a purification similar to the one that you describe," he wrote.

All this time, Dwyer had heard nothing about Richard's communication with Hellinga. After earning her doctorate in 2004, she had left Hellinga's lab in 2005 to pursue postdoctoral research in a different department. So she was not seriously concerned when Hellinga e-mailed her on the Labor Day holiday on 3 September last year, asking her to meet with him later in the week to discuss issues about NovoTIM. But Dwyer's new adviser, Donald McDonnell, a professor of pharmacology and cancer biology, advised her not to meet Hellinga alone; he felt she should go with someone who could advocate on her behalf. McDonnell arranged a meeting later that week at which he, Dwyer and Hellinga were joined by two other faculty members from the biochemistry department. And that's when Hellinga dropped the bombshell. "He said, 'I find it really hard to believe that you didn't make this up', and he kept saying that kind of statement over and over again," Dwyer says. "It was horrible."

Dwyer's adviser defended her, and she proclaimed her innocence. "I said, 'That's ridiculous, no, I didn't do that,'" she says. "What he was saying wasn't true."

A few weeks later, McDonnell, Hellinga, Dwyer and the head of the biochemistry department met again. Dwyer's husband, who is also a scientist, was there. Dwyer showed Hellinga the data from her lab notebooks that, she thought, exonerated her. But, she recalls, "he didn't want to look at any of that. It was just flat out my fault, and that was it." Hellinga remembers it differently. "That's not true," he says. "Of course I looked at the data. I also had people in my lab repeat the experiments," he says.

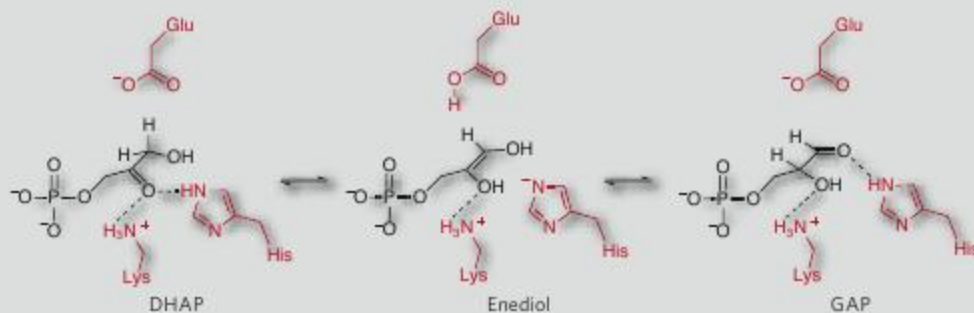
On 8 October, Hellinga wrote to Richard. "We have completed our repeat experiments on NovoTIM," he wrote. "I concur with your finding that the NovoTIM designs do not exhibit enzymatic activity, and that the reported activity is due to a contaminating activity which is very likely to be the endogenous, wild-type triose phosphate isomerase." The repeat negative control experiments, Hellinga wrote, had found "TIM activity in the wild-type [ribose-binding protein] preparations prepared by the step gradient elution method."

He added that the repeat experiments were done by three people, "but NOT Mary Dwyer, the author responsible for executing the experiments described in the *Science* paper, and responsible in large part for the negative control experiment in the *Journal of Molecular*



## HOW TIM WORKS

The enzyme called triose phosphate isomerase (TIM) catalyses one important step in sugar metabolism, and is found in most living organisms. The enzyme (active sites shown in red) allows cells to efficiently convert two sugars, called DHAP and GAP, into one another through the intermediate molecule enediol. Computer programs could enable researchers to design such efficient enzymes from scratch.



Biology paper." By naming Dwyer as the scientist primarily responsible for the experiments, Hellinga seemed to contradict his 30 July e-mail to Richard, in which he said "different individuals" had been involved. However, Hellinga clarified to *Nature* that his July e-mail was "slightly inaccurate"; at that time, Dwyer was the only person who had performed the negative controls, he says.

To Richard, Hellinga continued: "Dwyer has been contacted in an attempt to seek an explanation ... The matter has been referred to the Office of the Dean of the Medical School for further enquiries, which are now in progress."

A committee on research misconduct convened a formal inquiry hearing in December, at which Dwyer was asked to address the claims against her. On 4 February, she received a letter from Wesley Byerly, an associate dean in the medical school, clearing her of the allegation of falsifying and fabricating results.

## Culture of blame

But word about the inquiry had already spread, outraging chemists who felt it was wrong for a mentor to accuse a student of fraud. "It is reprehensible," says Frey. "It is up to the adviser to instruct the student, to guide the student to find out what problems exist with the data and their interpretation of it, and to show the student what the pitfalls are."

This February, both the *Science* and *Journal of Molecular Biology* papers were formally retracted. "The triose phosphate isomerase activity observed in our reported preparations can be attributed to a wild-type TIM impurity," stated the *Science* retraction; the other retraction was similar. Other chemists were surprised that Hellinga's lab had been fooled by a simple contamination problem. "It is a bush-league error not to purify your proteins well, especially in a paper like this," says Wallace Cleland of the University of Wisconsin-Madison.

Still, exactly what happened remains murky. On 10 March, *Science* published letters from

Richard and Kirsch listing issues they said were not resolved by the retractions. For instance, they wrote, the kinetic values Hellinga reported for NovoTIM are not the same as those of wild-type TIM, which is difficult to understand, given that all the activity in Hellinga's papers was supposed to have come from the wild-type enzyme.

Kirsch also raised questions about other experiments in the *Science* paper that "would make sense only if the design were successful". For instance, the paper reported that NovoTIMs could substitute for wild-type TIMs in *E. coli* that lack TIM enzymes. And a different test supposedly showed that mutated NovoTIMs became less active, just as DEZYMER had predicted. Neither of these results makes sense if the designed enzymes never worked.

Hellinga does not have explanations for the issues

Kirsch and Richard have raised. Dwyer thinks that the issues with protein expression and assay variability are partly to blame, and says that in retrospect, the apparent decreased activity of NovoTIM mutants was actually insignificant, once experimental error is taken into account. But no one has offered a clear answer for what went wrong. That is frustrating to Richard, who has spent considerable time and resources trying to get to the truth.

But thanks to Richard's work, another research team has been able to earn credit for the breakthrough Hellinga once claimed. In March, a team led by biochemist David Baker from the University of Washington in Seattle published two papers showing that computer programs could indeed be used to design working enzymes<sup>8,9</sup>.

Meanwhile, other scientists have questioned whether Hellinga himself should be investigated. Some point to a Duke policy that states that if an allegation of misconduct is found to be "baseless and malicious or reckless, the matter will be dealt with in accordance with existing university policies and mechanisms".

Hellinga says he has received no formal notification that he is under investigation. Duke would not comment specifically, saying only: "We are aware the retraction by Dr Hellinga has generated considerable debate in the scientific community. Duke continues to follow this debate and is evaluating various points that are being raised."

Asked whether he would have done anything differently in the NovoTIM experiments, Hellinga says, "I would like to not have the problem that we encountered." When asked whether the lab moved too quickly, he says: "Given how we understood things to be at the time, no. Obviously if we had known things had gone wrong, we wouldn't have moved forward with the speed we did."



John Richard flagged issues with potential contamination.

As for Dwyer, she still feels rattled by the experience. "I feel incredibly guilty that I didn't catch it, but I didn't, and I just have to live with that. It's been really hard," she says. She is trying to move forwards with her life and career, she says, and is working in a new lab in a new field — endocrinology — with McDonnell. But sometimes, Dwyer says, she thinks back to the people who tried to steer her away from Hellinga's lab so many years ago. And she wonders how different things might have been if she had heeded their advice.

"Everybody gets warned, but nobody listens," she says. "Maybe now they will."

**Erika Check Hayden is a senior reporter in Nature's San Francisco office.**

- Hellinga, H. W. & Evans, P. R. *Nature* **327**, 437–439 (1987).
- Hellinga, H. W. & Richards, F. M. *J. Mol. Biol.* **222**, 763–785 (1991).
- Bolon, D. N. & Mayo, S. L. *Proc. Natl Acad. Sci. USA* **98**, 14274–14279 (2001).
- Dwyer, M. A., Looger, L. L. & Hellinga, H. W. *Science* **304**, 1967–1971 (2004).
- O'Donoghue, A. C., Amyes, T. L. & Richard, J. P. *Biochemistry* **44**, 2610–2621 (2005).
- O'Donoghue, A. C., Amyes, T. L. & Richard, J. P. *Biochemistry* **44**, 2622–2631 (2005).
- Allert, M., Dwyer, M. A. & Hellinga, H. W. *J. Mol. Biol.* **366**, 945–953 (2007).
- Jiang, L. *et al. Science* **319**, 1387–1391 (2008).
- Röthlisberger, D. *et al. Nature* **453**, 190–195 (2008).

See Editorial, page 258.



# CORRESPONDENCE

## Deforestation: call for justice, not militarization

SIR — As a recent policy adviser to the United Nations in a programme intended to address environmental threats in the Amazon, I would caution readers of your Special Report 'Brazil goes to war against logging' (*Nature* **452**, 134–135; 2008). Although this analysis of Brazil's "military-style crackdown on deforestation" is valuable, it perpetuates some perilous assumptions.

You aptly note that mining on the fringes is among the problems, but I would question some of the solutions you suggest. Urging international donors to help national governments such as Brazil to fund 'traditional' policies for enforcement neglects the reality that local governance agents, not national enforcers, can be more appropriate custodians; foreign donors frequently make this mistake. Also, there needs to be greater global awareness of the unwanted consequences of the so-called traditional policies of eco-militarization, which have wrought pervasive problems not only in the Amazon but worldwide.

Police interference in indigenous people's lives can have devastating effects, often leading to more environmental harm rather than less. My own experience with peasants who mine illegally in conservation areas suggests that resentment of the government drives illegality and marginalization. For example, indigenous groups in the Amazon captured more than 600 hostages in the past two years, protesting against injustices with regards to mining rights. One tribe took a UN worker and four others hostage a few months ago, angry that the government and international agencies ignored their resource claims; the government sent in police troops instead. That tribe was seeking the rights to mine its land — but only a multinational corporation was granted those

rights. Brazil has millions of peasant ('artisanal') miners, but licences only very rarely go to poorer communities.

Some 337 large-scale commercial mining permits violate Brazilian law, falling within 28 federal conservation areas; another 5,283 pending applications also impinge on federal conservation areas. Environmental violations are pervasive in both legal and illegal scenarios — and the concept of legality itself is blurred.

It is time to politicize the Amazon, not militarize it. Conservation advocacies should focus on socio-economic empowerment of vulnerable populations. Military enforcement should be a rare last resort.

It is time to listen to the poorest 'illegal' people and let them shape policy. Creating new mega-reserves, even if appropriate, is rarely an effective solution. Combating deforestation requires helping workers to find livelihoods while participating equitably in the intricate politics of nature.

**Samuel J. Spiegel** Geography Department, University of Cambridge, Downing Place, Cambridge CB2 3EN, UK

## Deforestation: damage from dams adds to emissions

SIR — In your Special Report 'Brazil goes to war against logging' (*Nature* **452**, 134–135; 2008), you argue that the Brazilian government is willing to stop logging operations on the Amazon and to push down Brazil's greenhouse-gas emissions. But the Brazilian government still plans to build several hydroelectric power plants on the Amazon (see <http://tinyurl.com/5hb6st>), which could increase both deforestation and greenhouse-gas emissions.

Dams flood thousands of hectares of forest and threaten many aspects of the Amazon basin. Some are already being

built in the Amazon headwaters and will interfere with the migratory route of economically and ecologically important fish. The previous construction of such barrages was a huge mistake, as sediment transportation to the lowlands (where soil is poor) was blocked and the excess of dead trees in the reservoir caused high water eutrophication.

Dams have also been implicated in methane and carbon dioxide emissions, in part because of organic-matter decomposition in the reservoirs. These emissions from artificial lakes should be considered along with Amazon logging and burning in calculations of Brazil's total greenhouse-gas emissions.

The Brazilian government needs to look at some proposed solutions, such as renewing non-productive 30-year-old dams (already built next to energy-consuming regions, but they would keep methane emissions at today's levels) or imposing a solar-energy policy that could reduce greenhouse-gas emissions if dams were decommissioned. Unfortunately, it does not look as though this will happen.

**André Frainer Barbosa** Post-Graduation Program on Ecology, Universidade Federal do Rio Grande do Sul, CEP: 91501-970, Porto Alegre, Brazil

## Hype around nanotubes creates unrealistic hopes

SIR — In your News story 'Stem-cell claim gets cold reception' (*Nature* **452**, 132; 2008), you report on an announcement that nanotubes have been used to reprogramme adult human cells. As you say, the result was met with scepticism because it was proclaimed at an investors' meeting, rather than in a peer-reviewed publication. There are also questions about the nanotechnological aspects of the work.

Carbon nanotubes are capable of penetrating cell membranes to facilitate cytoplasmic delivery, but their application in medicine is still in its infancy. Much more basic research is needed before deterministic assertions — such as those contained in the company announcement — can be made about the way carbon nanotubes act biologically and pharmacologically. Furthermore, no preclinical therapeutic efficacy data exist yet for any disease using a carbon nanotube construct.

The announcement was unclear about the exact nature of the studies carried out and the promised capabilities of both the carbon nanotubes and stem cells. For example, what was actually delivered by the carbon nanotubes? Proteins, DNA, small interfering RNA, or all three? Dramatically different delivery-system specifications are necessary for successful intracellular delivery and guaranteed biological activity for each of these, making the notion that 'carbon nanotubes deliver all' too simplistic.

The field of medical applications of carbon nanotubes and nanotechnology in general is particularly prone to hype. Claims that nanomaterials can contribute to new treatments of devastating diseases have led to confusion and misrepresentation, risking discreditation of this exciting area of research. We urge caution and patience, and call for sustained investment in basic research and systematic investigation by multidisciplinary teams, to keep expectations realistic.

**Kostas Kostarelos** Nanomedicine Laboratory, Centre for Drug Delivery Research, The School of Pharmacy, University of London, 29–39 Brunswick Square, London WC1N 1AX, UK  
**Alberto Bianco** CNRS, Institut de Biologie Moléculaire et Cellulaire, Laboratoire d'Immunologie et Chimie Thérapeutiques, 15 rue René Descartes, 67000 Strasbourg, France  
**Maurizio Prato** Dipartimento di Scienze Farmaceutiche, Università di Trieste, Piazzale Europa 1, 34127 Trieste, Italy



**"Those years were the darkest of my professional career."**

Tony Fauci on HIV — see page 289



## A case of genetic counselling for Dr Watson

SIR — Your recent publication of James Watson's sequenced genome (D. A. Wheeler *et al.* *Nature* **452**, 872–876; 2008) makes for fascinating reading — that is, for anyone who has enough knowledge of genetics and biology to interpret it. Peel away the single nucleotide polymorphisms (SNPs), the copy-number variations and the indels, and the core message seems, well, empty. As Maynard Olson

admits in the accompanying News & Views (*Nature* **452**, 819–820; 2008): "If Watson took his sequence to a genetic counsellor, there would be little to discuss."

As a genetic counsellor, I wonder how a session with Watson would stack up against one with a non-scientific member of the general population? The table offers some insight.

Where's the challenge in providing genetic counselling to 'Dr W'?

**Myra I. Roche Medical School Wing E, CB # 7487, University of North Carolina, Chapel Hill, North Carolina 27599-7487, USA**

Dr Watson	Lay patient
Discovered the structure of the gene	Not exactly sure what a gene is
Thinks the \$1 million cost is a good deal	Worried about the cost of a consultation
Sequence ready in 2 months	Standard test result ready in 2 months
Brings in sequence data on a hard drive	Brings in records about sinus infection
Can give fully informed consent	Can sign a consent form
Released his sequence to the public	Doesn't want mother to know
Chose to have Apo-E sequence redacted	Expects to learn blood type
Understands the limited clinical relevance	Hopes we can fix any broken genes
Shares 1.68 million SNPs with Craig Venter	Googles SNPs to find out who they are
Concerned about not being cloned	Concerned about being cloned

## The public needs to know social benefits of vaccination

SIR — You may have been inadvertently off message about vaccination policy in your Editorial 'On message, off target' (*Nature* **452**, 128; 2008) — for what you did not say. Centuries of experience have proved the importance of societies (not doctors, parents or vaccine recipients) in making decisions on whether to vaccinate or not. Having drafted the US national vaccine programme legislation that was passed in 1986, and directed the national vaccine programme in 1993–94, I can say that recent UK experience impressively illustrates the strategic distinction.

After an unexpected rise of pertussis (whooping cough) cases in the 1990s, Britain revamped its system to reward general practitioners for fully immunizing all children on their lists, paying a bonus to those who reached 90% then, a few years later, 95%. This system did not exclude from the denominator the small number of patients with good medical reasons against receiving vaccine.

Why the societal strategy? No vaccine is fully effective in everyone who receives it. Only 'herd immunity' can protect those in whom a vaccine does not stimulate protective immunity. Population health therefore improves where policies support full coverage of the population. Such policies create a culture in which health authorities educate doctors, parents and patients about the societal goals as well as individual contraindications.

Good immunization policies balance the risks of vaccines against the benefits. All vaccines, usually given to healthy people, pose some small risk. Good medical reasons do occasionally exist for some people to avoid some vaccines. Britain understood this, and designed its policy to avoid incentives for doctors to vaccinate in such situations.

Although vaccination remains voluntary, no programme is based on individual preference. Public education about vaccination is important, particularly to heighten understanding of the basis for all aspects of public policy to maximize protection of human health.

**Anthony Robbins Journal of Public Health Policy, 213 West Canton Street, Boston, Massachusetts 02116, USA**

## Public support never has guaranteed good work

SIR — Sadly, the implication in your Editorial 'Broken promises' (*Nature* **452**, 503; 2008), that good science depends on good ideas and not good intentions, is unlikely to make much difference to politicians anxious for public favour or to scientists anxious for money.

As you declared more than 15 years ago (*Nature* **360**, 13; 1992), "the increasingly crass politicization of biomedical research looms as something of a menace, for it presupposes first that more money for more scientists is a way to medical salvation, and second that well-meaning groups of citizen-activists and professional lobbyists have a scientifically useful role in deciding where research money should be directed. There is, alas, inadequate evidence to support either proposition." Technology can be bought, but science depends first on ideas and luck, to which the link is more imprecise.

In that News & Views article, you give four examples where politics threatened to derail science. One of them — five years before the Clinton promise mentioned in the Editorial (it is fascinating just how many breakthroughs are promised "within a decade") — was of the US National Institutes of Health and Food and Drug Administration trying to stop lobbyist-driven testing of an unproven AIDS vaccine.

**Neville W. Goodman Bristol, UK**



## BOOKS &amp; ARTS

## Big problems, big decisions

Sustainable solutions to worldwide crises such as overpopulation and climate change need regulating by global bodies, but whose views should these organizations represent?

Supranational organizations are necessary to monitor the welfare of our planet. They alert us to impending crises, develop policies to avert dangers and induce governments to respond. But it is easy to lose sight of who these organizations represent.

Two books — *Fatal Misconception* by historian Matthew Connelly and *Common Wealth* by economist Jeffrey Sachs — highlight the importance of knowing who speaks for whom. Connelly describes how the international family-planning movement responded to concerns of an imminent population explosion. Sachs assesses our stewardship of this crowded planet and suggests how we should conduct its affairs.

In *Fatal Misconception*, Connelly begins his story in the early 1950s, when public-health advances reduced mortality among the impoverished populations of newly independent nations. Demographers, birth-control activists and eugenicists began to worry that population growth might outstrip the global food supply. Militating against limits to population growth were the Vatican and a US law (the Comstock Act of 1873) that prohibited birth control by married couples and the sale of contraceptive devices. In 1952, the Vatican secured a veto preventing any involvement of the fledgling World Health Organization in population matters. Nonetheless, a worldwide movement emerged that was initially dedicated narrowly to population control, but which eventually came to see questions of fertility as part of the developmental progress of nations.

Two international conferences, held four decades apart, encapsulate this change. The first was convened by US philanthropist John D. Rockefeller III in 1952 at the still-segregated Williamsburg Inn in Virginia. The delegates (mostly white men) created two organizations dedicated to the revolutionary idea of securing a global reduction in population growth: the Population Council and the International Planned Parenthood Federation (IPPF). To be credible, the organizations sought international involvement. A window of opportunity materialized



Family matters: a clinic doctor advises women on family planning and contraception in India.

in 1952 when Indian president Jawaharlal Nehru announced the world's first national population policy for India.

The second meeting, on population and development, was sponsored by the United Nations and held in Cairo in 1994. The delegates endorsed a radical new policy asserting the right of women to control their own fertility, together with improvements in maternal health and female literacy, as the best route to stabilizing the population. Opposition was slight, and came from an alliance of conservative Muslim and Catholic states.

The mass birth-control campaigns of the 1950s were improvised affairs that used inadequately tested devices and procedures. Connelly leads us adroitly through extraordinary and complex incidents, particularly the interventions in India. The first Indian family-planning campaign at Khanna, Punjab, was initiated by a team from Harvard University, Cambridge, Massachusetts, and failed disastrously; subjects agreed to participate but would not use the contraceptives. Threatened by famine, India was blackmailed by US president

Lyndon B. Johnson's administration into accelerating its population programme by the promise of food aid. In an atmosphere bordering on panic, two coercive family-planning campaigns in India spawned a legacy of bitterness towards experts and politicians. Meanwhile, fears of communism fuelled clandestine support from the US government for the IPPF's worldwide birth-control campaigns, conveniently out of sight of Congress. India later expelled the family-planning consultants in exasperation over US partiality for Pakistan. Nevertheless, the US Agency for International Development (USAID), working indirectly through a United Nations agency, inundated the developing world with contraceptive devices. In the 1980s, Ronald Reagan's administration withdrew US support for UN population policy.

By 1990, the average family size had fallen from six or more children to around three almost everywhere apart from sub-Saharan Africa. Connelly believes that intense birth-control campaigns contributed little to this decline (except in China). Fertility was already falling before the campaigns started, owing to social changes, but was probably assisted by the broad availability of contraceptives, including those distributed through USAID.

Connelly suggests that the "fatal misconception" of the population movement was its

**Fatal Misconception: The Struggle to Control World Population**  
by Matthew Connelly

Harvard University Press: 2008. 544 pp.  
\$35.00, £22.95

**Common Wealth: Economics for a Crowded Planet**  
by Jeffrey D. Sachs

Penguin/Allen Lane: 2008. 400 pp. \$27.95, £22



conviction that it knew the best interests of poor agrarian people. They needed large families in case some of their children died prematurely to ensure provision of care in old age and a labour force.

Continuing the theme of global stewardship, Sachs' book *Common Wealth* is a manifesto for securing a bright future for Earth. A key plank in Sachs' platform is the stabilization of world population to minimize the tension between industrialized and developing nations. The profligacy of industrialized life is the most severe threat to the environment, and can only get worse as China and India add to the burden. Some of the most fragile habitats on Earth are being degraded rapidly by population pressure.

Just as everyone in democratic societies has a stake in their communal arrangements, Sachs argues that a global society must evolve to meet the ecological, demographic and economic challenges we face together. Reaching a shared prosperity requires the sustainable use of energy, land and resources, stabilization of population at around eight billion by 2050 and the elimination of extreme poverty by 2025. Arguing that solutions are possible, Sachs points to our success in negotiating other potential subsistence crises through scientific discovery. New technologies and reduced exploitation of natural resources may suffice to avert future catastrophes.

Sachs presents sobering statistics: we use 45% of Earth's photosynthetic potential; 60% of the run-off water and, of all the

atmospheric nitrogen fixed, 55% is obtained through the Haber-Bosch chemical process. Fertilizers from this process underpinned production of 80% of cereals consumed in the twentieth century; without it, the world's population would be only one-quarter of what it is at present. During the past 50 years, carbon dioxide levels have increased by 25% and may double by 2050. If so, the global temperature is predicted to rise by 3 °C, drastically increasing sea levels and reducing the availability of fresh water. At certain critical temperature thresholds, events will occur that push temperatures even higher, such as the release of methane sequestered in the frozen tundra or in semi-solid methane hydrates deep in the oceans.

People living in ecologically fragile regions (particularly in sub-Saharan Africa), areas already hampered by deforestation, overgrazing and overuse of underground aquifers, may find subsistence farming impossible if fresh water becomes scarce. The 10% of people who live within 10 metres of sea level will become vulnerable to increasingly common storms and floods. Subtropical regions that were free of mosquitoes may suddenly become malarious. Sachs fears that such calamities will provoke migration and conflict. In societies with excessive birth rates, a 'youth bulge' may outnumber adult authority figures, creating social instability. High birthrates also hinder economic

development, trapping families in poverty by absorbing any rise in productivity.

What can be done? Ever the optimist, Sachs believes we can pull back from the brink if we cooperate to reduce CO<sub>2</sub> emissions. He hopes for technology to sequester such emissions, but reiterates the widely discussed need for an end to deforestation, a switch to non-fossil fuels, more efficient power consumption, the use of electric hybrid vehicles and more energy-efficient buildings.

Sitting uneasily in this exploration of the ecological impact of the industrialized world is a discussion of what should be done to help the one-sixth of the world's population trapped in extreme poverty. Sachs argues that

they can escape the poverty trap only if agricultural yields and education improve, and if modern infrastructure, sanitation and health care are available. Sachs sees the UN Millennium Development Goals as a vital framework for reducing poverty. Ruefully, he reminds us that even this altruistic scheme has its enemies; for example, the US ambassador to the United Nations tried to expunge the concept from the agenda of the 2005 UN summit.

With great faith in international organizations, Sachs points to the control of chlorofluorocarbons, leaded petrol and acid rain as remarkable tributes to their value. For its breadth, the book is mysteriously silent on the role of the International Monetary Fund (IMF) and the World Bank, institutions created to lift nations out of poverty, and which should hold the key to solving problems of development. The economics Nobel laureate Joseph Stiglitz, writing in his 2003 book *Globalization and its Discontents* (W. W. Norton), damned the IMF as having demonstrably damaged the poor of the developing world since the era of Reaganomics. Omission of this topic seems to reduce the value of *Common Wealth*.

*Fatal Misconception* describes a historic clash of opposed interest groups wrestling to impose their own population policies on the developing world. *Common Wealth* is an audacious first draft of a plan for sustainable development of the planet, in which population is one issue. If global institutions implement Sachs' plans, there will be winners and losers; we should be sure to ask who these organizations represent.

**Michael Sargent** is a developmental biologist at the National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. He is author of *Biomedicine and the Human Condition: Challenges, Risks and Rewards*.

**"A global society must evolve to meet the ecological, demographic and economic challenges we face together."**



In 1960s India, elephants were trained to distribute leaflets about birth control to villagers.



# Enshrining the right to live or die

Developments in the life sciences inevitably raise complex legal questions, such as when a human life begins and ends, and how much weight should be given to individual preferences about how and when we reproduce, or die. The law cannot avoid getting involved in these finely balanced bioethical issues, but courts and legislators are sometimes spectacularly ill-equipped to deal with them.

Take the sad case of Terri Schiavo, described in Barry Schaller's riveting book *Understanding Bioethics and the Law*, which describes a series of confrontations between US law and some difficult bioethical dilemmas. Eight years after Terri lost consciousness and fell into a permanent vegetative state (PVS), her husband Michael sought court authorization for the removal of Terri's feeding tube. By this time, Michael's relationship with Terri's parents, Robert and Mary Schindler, had broken down, and the Schindler family objected to the removal of the tube. Michael and the Schindlers then engaged in virtually continuous litigation for the next five years during which the tube was removed, only to be reinstated two days later. The family were joined in this unedifying legal spectacle by various religious interest groups, advocacy organizations representing people with disabilities, and by the Florida House of Representatives. With the support of governor Jeb Bush, the Florida legislature passed Terri's Law, a statute preventing the withdrawal of nutrition and hydration from a patient "who met the unique description of Terri Schiavo". After further legal challenge, Terri's Law was eventually ruled unconstitutional by Florida's Supreme Court.

The action then shifted to the federal level, and a bill to nullify the Florida court's decision was introduced in the House of Representatives and in the Senate. President George W. Bush returned early from his holiday to sign the Senate bill at 1 a.m. on 21 March 2005. All this legislative activity did not derail the court process. The US Supreme Court refused to hear an appeal to overturn the lower court's decision that Terri's tube could be removed. Terri eventually died on 31 March 2005.

Fortunately, the Terri Schiavo case is exceptional: most end-of-life decisions are resolved privately, through discussions between family

members and clinicians. Although there are many differences between Schaller's book and *Easeful Death* — Mary Warnock and Elisabeth Macdonald's elegant and thoughtful discussion of the case for assisted dying — the authors agree that private, consensual resolution of complex individual cases is almost always preferable to exposing them to the harsh glare of litigation and legislation. Warnock and Macdonald, for example, seem to regret the ending of the 'humane practice' through which people could ask their 'retired medical friends' for help in storing up "a collection of drugs for future use if they found life intolerable".

The Terri Schiavo case, together with that of Dianne Pretty — a British woman with motor neurone disease who wanted to be helped to die — suggest that litigation may not resolve conflict. But at least court proceedings do focus on the individual. Schaller contrasts the US courts, which stuck to the question of whether Terri would have wanted to be kept alive in a PVS, with the "executive and legislative branches" who "disgraced themselves by pushing the boundaries of law and decency in their self-aggrandizing efforts". Although this harsh criticism seems fitting given the political exploitation of this "private ethical dilemma about a young woman's sad decline", it points to a wider problem in trying to use legislation to resolve complex bioethical issues.

Although the adversarial process has its disadvantages, the courts can at least attempt to deliver individualized justice, but legislation must apply to all. Translating compassionate responses to the suffering of someone such as Dianne into law is difficult precisely because legislation cannot just apply to one individual's situation: it must apply to everyone. Warnock and Macdonald quote Robin Gill, a member of the British Medical Association's medical ethics committee, who made a submission for Dianne on compassionate grounds. Her case, in Gill's view, "represented a very, very strong case indeed for voluntary euthanasia". If it was just about her, and no one else, he would be in favour of giving her access to assisted dying. But legislation cannot be personalized in this way, and Gill concluded "that more people will be made more



Terri Schiavo's case challenged bioethics law.

vulnerable if we change the law in favour of legalizing assisted dying".

Once an issue is before the legislature, the suffering of the individual, and the possibility of a compassionate and individualized response to that suffering, is lost. Someone such as Dianne has to die in a distressing manner, not because we think that she is vulnerable and incapable of making a considered, informed and voluntary decision to end her life, but because we think that other people might be vulnerable. Is this fair? Many would say no, and if we support Dianne's right to assisted dying, then we should not expect her to suffer for the greater good of society. Warnock and Macdonald accept that it would be hard to draft a law that would always succeed in protecting the vulnerable, but they conclude that, for compassionate reasons, we should at least try.

The legislative process offers an opportunity for special-interest groups to intervene. In the Terri Schiavo case, the concern of many advocacy groups was not the dignity and prior wishes of Terri herself, but the effect that her case might have on other people. If Terri were allowed to die, they asked, how would this affect other Americans with disabilities? These powerful advocacy groups formed an alliance with the even more powerful conservative religious lobby, leading to what Schaller describes as a "media mega-event".

The cases of Terri Schiavo and Dianne Pretty raise different issues. Terri never contemplated ending up in a PVS, and her views

**Understanding Bioethics and the Law: The Promises and Perils of the Brave New World of Biotechnology**  
by Barry R. Schaller  
Praeger Publishers: 2008. 241 pp.  
\$49.95, £27.95

**Easeful Death: Is There a Case for Assisted Dying?**  
by Mary Warnock and Elisabeth Macdonald  
Oxford University Press: 2008.  
155 pp. \$26



on the withdrawal of life-prolonging treatment were unknown, whereas Dianne was intelligent and competent, and had clearly articulated views on how she wanted her life to end. It is odd, perhaps, given that Terri would know nothing of the court's decision, that the outcome hinged on what her husband and parents claimed her wishes would have been, as a result of the so-called substituted judgement test. By contrast, the views of Dianne, who went on to suffer greatly before

she died, were largely irrelevant and trumped by the possibility of harm to others.

The law moves in ways that are mysterious to many. Both these fascinating books, although dealing with different jurisdictions and issues, contribute to our understanding of the most important ethical challenges that lawyers will face in the coming decades.

**Emily Jackson** is professor of law at the London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

cheap and rarely happens overnight", Monmonier tells the story of the ambitious International Map of the World project. Following a German geographer's proposal at an international conference in 1891, cartographers from the industrialized nations agreed to produce a series of maps with uniform projection, symbols, style, scale and technique, which would adequately describe all of the mountains, rivers and key coastal features of the world. When complete, the mosaic of assembled 'millionth maps' — drawn to a scale of 1:1,000,000 — would have covered more than 185 square metres (half the size of a basketball court) at a cost of £100,000 (US\$200,000), a significant sum at the time. Some maps were eventually produced but, with most nations reluctant to pay their share and with two intervening world wars, this grand project ended in a whimper in the second half of the twentieth century.

The failure of the International Map of the World project can, to some extent, be explained by emerging technologies that made the old standards obsolete. Some advances have revolutionized data collection and display, and some (such as intercontinental ballistic missiles) expanded the military's need for maps that were increasingly accurate and sophisticated. Monmonier has good knowledge of cartographic technologies. In a section on overhead imaging, he introduces the problem of consolidating images taken from different angles and discusses techniques for remote sensing. In another chapter he surveys the types of electronic chart that are now available.

Today, the mapping of coastlines has environmental and political implications. Proposing that cartographers can document global climate

## Charting the water's edge

**Coast Lines: How Mapmakers Frame the World and Chart Environmental Change**  
by Mark Monmonier

University of Chicago Press: 2008. 224 pp. \$25

Coastal margins are drawn as precise lines on maps, yet anyone who has walked on a beach knows that coastlines are unstable. Cliffs, rocks and sands change under the influence of tides, storms, tectonic movements, global climate change and other natural and artificial phenomena.

Mark Monmonier, professor of geography at the Maxwell School of Syracuse University in New York, seeks to inform the public about how cartography and society intersect. He wishes us to look closely at maps, to recognize which features are shown or missing, and understand why. In *Coast Lines*, he offers an assortment of eclectic and fascinating information about how coastlines have been defined, determined and depicted, focusing on the United States in the twentieth century.

Different maps and charts of the same coastal area show different cartographic coastlines. Monmonier calls our attention to four types, explaining that each is a human construct designed to serve a specific purpose, and the result of many observations and assumptions (the latter sometimes gaining the upper hand). One cartographic coastline is the high-water line visible from offshore. Another, introduced in the nineteenth century to aid safe navigation, is the low-water line. Two are more recent: storm surge lines are designed mainly for evacuation planning and flood insurance, and inundation lines describe the plausible effects of changing geological and meteorological conditions.

Monmonier discusses the international boundaries of territorial waters. These comprise the region beyond a nation's coastline, however defined, in which that nation has sovereignty, where it can "enforce laws, levy taxes, and

exclude foreign vessels not pursuing expeditious 'innocent passage' along the coast or into a port". Under pressure from military and economic quarters, territorial waters have, in recent decades, been extended from the traditional 5.5 kilometres to 22 kilometres from shore.

*Coast Lines* introduces the exclusive economic zone (EEZ) in which maritime nations can "manage fisheries, mine the sea bed, and extract oil and natural gas" but cannot exclude foreign vessels. The EEZ extends 370 kilometres from shore, except where it encounters the EEZ of another nation, in which case the boundary is drawn equidistant between the two coastlines. Islands can affect where the line is drawn. For instance, because of the position of Key West and the Dry Tortuga islands off the southwestern coast of Florida, the compromise boundary between the United States and Cuba is much farther south than would be expected.

To show that "good cartography is seldom



Satellite images help map-makers to chart changes in coastal margins, such as shifting sands from Louisiana to Alabama.

NASA/JPL-CALTECH



on the withdrawal of life-prolonging treatment were unknown, whereas Dianne was intelligent and competent, and had clearly articulated views on how she wanted her life to end. It is odd, perhaps, given that Terri would know nothing of the court's decision, that the outcome hinged on what her husband and parents claimed her wishes would have been, as a result of the so-called substituted judgement test. By contrast, the views of Dianne, who went on to suffer greatly before

she died, were largely irrelevant and trumped by the possibility of harm to others.

The law moves in ways that are mysterious to many. Both these fascinating books, although dealing with different jurisdictions and issues, contribute to our understanding of the most important ethical challenges that lawyers will face in the coming decades.

**Emily Jackson** is professor of law at the London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

cheap and rarely happens overnight", Monmonier tells the story of the ambitious International Map of the World project. Following a German geographer's proposal at an international conference in 1891, cartographers from the industrialized nations agreed to produce a series of maps with uniform projection, symbols, style, scale and technique, which would adequately describe all of the mountains, rivers and key coastal features of the world. When complete, the mosaic of assembled 'millionth maps' — drawn to a scale of 1:1,000,000 — would have covered more than 185 square metres (half the size of a basketball court) at a cost of £100,000 (US\$200,000), a significant sum at the time. Some maps were eventually produced but, with most nations reluctant to pay their share and with two intervening world wars, this grand project ended in a whimper in the second half of the twentieth century.

The failure of the International Map of the World project can, to some extent, be explained by emerging technologies that made the old standards obsolete. Some advances have revolutionized data collection and display, and some (such as intercontinental ballistic missiles) expanded the military's need for maps that were increasingly accurate and sophisticated. Monmonier has good knowledge of cartographic technologies. In a section on overhead imaging, he introduces the problem of consolidating images taken from different angles and discusses techniques for remote sensing. In another chapter he surveys the types of electronic chart that are now available.

Today, the mapping of coastlines has environmental and political implications. Proposing that cartographers can document global climate

## Charting the water's edge

**Coast Lines: How Mapmakers Frame the World and Chart Environmental Change**  
by Mark Monmonier

University of Chicago Press: 2008. 224 pp. \$25

Coastal margins are drawn as precise lines on maps, yet anyone who has walked on a beach knows that coastlines are unstable. Cliffs, rocks and sands change under the influence of tides, storms, tectonic movements, global climate change and other natural and artificial phenomena.

Mark Monmonier, professor of geography at the Maxwell School of Syracuse University in New York, seeks to inform the public about how cartography and society intersect. He wishes us to look closely at maps, to recognize which features are shown or missing, and understand why. In *Coast Lines*, he offers an assortment of eclectic and fascinating information about how coastlines have been defined, determined and depicted, focusing on the United States in the twentieth century.

Different maps and charts of the same coastal area show different cartographic coastlines. Monmonier calls our attention to four types, explaining that each is a human construct designed to serve a specific purpose, and the result of many observations and assumptions (the latter sometimes gaining the upper hand). One cartographic coastline is the high-water line visible from offshore. Another, introduced in the nineteenth century to aid safe navigation, is the low-water line. Two are more recent: storm surge lines are designed mainly for evacuation planning and flood insurance, and inundation lines describe the plausible effects of changing geological and meteorological conditions.

Monmonier discusses the international boundaries of territorial waters. These comprise the region beyond a nation's coastline, however defined, in which that nation has sovereignty, where it can "enforce laws, levy taxes, and

exclude foreign vessels not pursuing expeditious 'innocent passage' along the coast or into a port". Under pressure from military and economic quarters, territorial waters have, in recent decades, been extended from the traditional 5.5 kilometres to 22 kilometres from shore.

*Coast Lines* introduces the exclusive economic zone (EEZ) in which maritime nations can "manage fisheries, mine the sea bed, and extract oil and natural gas" but cannot exclude foreign vessels. The EEZ extends 370 kilometres from shore, except where it encounters the EEZ of another nation, in which case the boundary is drawn equidistant between the two coastlines. Islands can affect where the line is drawn. For instance, because of the position of Key West and the Dry Tortuga islands off the southwestern coast of Florida, the compromise boundary between the United States and Cuba is much farther south than would be expected.

To show that "good cartography is seldom



Satellite images help map-makers to chart changes in coastal margins, such as shifting sands from Louisiana to Alabama.

NASA/JPL-CALTECH



change, Monmonier juxtaposes two charts of the Outer Banks of North Carolina that show how the coastline in this area has receded over the past 125 years. Turning to predictive cartography — that is, maps and charts that indicate the probable future effect of rising seas and coastal subsidence — he admits that their foundations incorporate “multiple sources of uncertainty”. Monmonier acknowledges the sensitive political

nature of materials of this type. For example, although he has seen the latest maps of tidal marshlands prepared by the US Environmental Protection Agency, he is prevented from discussing them before their authorized release.

By highlighting a selection of topics, *Coast Lines* may succeed in its goal of getting the public to think about what maps show and why. However, Monmonier cautions

prospective navigators to be aware that modern global-positioning technology enables them to know their latitudes and longitudes with greater precision than the scientists who made their maps in the first place.

**Deborah Jean Warner** is curator of the physical sciences collections at the National Museum of American History, Smithsonian Institution, Washington DC 20013-7012, USA.

## Changing perceptions of light

**Take Your Time: Olafur Eliasson**  
MoMA New York  
Until 30 June 2008.

The vast foyer of New York's Museum of Modern Art (MoMA) is being dive-bombed by a small electric fan. Suspended from the ceiling by wire, it propels itself on an erratic course, buzzing over the spectators like a giant insect. Rather more calming is the title work of the current exhibition, *Take Your Time*, by the Danish-Icelandic artist Olafur Eliasson. It features a 12-metre-wide mirrored disc, positioned off-kilter so that as it slowly revolves it distorts the lines of the room. Hypnotic and unsettling, it is best experienced while lying on the floor.

Eliasson is an experimental artist who engages with science and the natural world. In 2003, he filled the immense Turbine Hall of London's Tate Modern gallery with a giant artificial sun made of 200 yellow lamps, mirrors and mist. Its warm glow induced crowds to recline on the gallery floor as if basking on a beach.

Fascinated by our perception of visual effects, Eliasson co-hosted a conference at Columbia University, New York, on ‘The Colors of the Brain’ to coincide with his 38-piece exhibition. The artist led his audience through an experiment designed to induce optical after-images, asking them to gaze into the bright lights overhead, then stare for a minute at a bottle placed on a chair, and then look at a blank screen. “The pink colour you see is the colour of blood in the eyes,” he explained. Both scientists and art historians took part: neuroscientist David Eagleman discussed optical illusions and synaesthesia, and the string theorist Brian Greene made a valiant effort to explain his own abstract work.

“Science,” Greene asserted, “reveals a reality different from what we perceive.” Eliasson similarly questions our precarious notion of what is external and what is created in the mind of the perceiver. His *360° Room For All Colours*

— on view at MoMA — is a backlit drum of projected light that shines bright pink, bleeds to white, then turns electric blue. The colours seem to loiter in the mind's eye even after they have gone. Another room is starkly lit by flashes of different tones of white (some cool, others warm), the subtle changes in hue revealing the white cube gallery space to be less of a neutral background than it might initially appear.

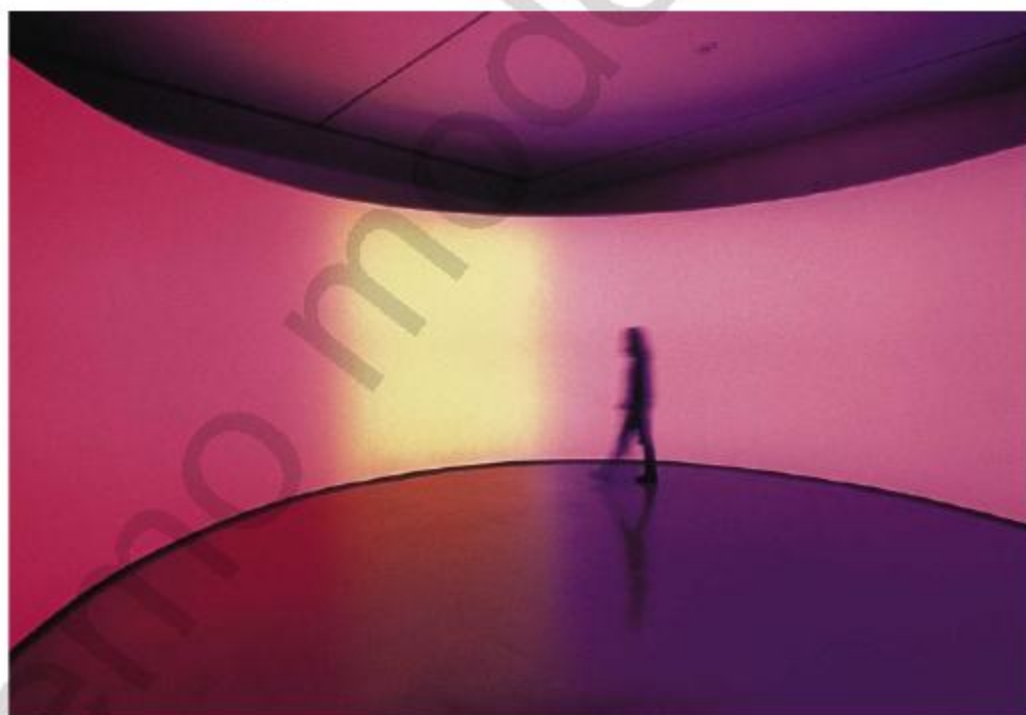
Two rooms display Eliasson's maquettes and architectural models. Contrasting with the minimalist museum installations, these give us a glimpse inside Eliasson's working mind — the rooms are a crazy mix of revolving spheres, kaleidoscopes, geodesic domes and chandeliers made of shards of mirrored glass.

Much of Eliasson's work deals with the intersection of natural and mechanical elements. In *Your Strange Certainty Still Kept*, an aggressive strobe light freeze-frames water droplets as they sprinkle into a trough from a perforated hose. The stationary drops sparkle

and glitter, as if suspended in space. In *Beauty*, gallery viewers may walk through a fiery, pulsating rainbow curtain of fine mist that is illuminated by a single spotlight. In *Reversed Waterfall*, a makeshift stack of metal bins, pumps and jets are assembled to force the water to flow uphill.

In late June, and on a different scale, the city's Public Art Fund will launch Eliasson's US\$15-million New York City Waterfalls project, a series of four enormous and noisy man-made torrents measuring between 27 and 37 metres high. They will be located under the Brooklyn Bridge and at other sites along the East River. It will be the largest art installation in the city since Christo filled Central Park with his saffron-coloured Gates in 2005 and, for the few days that this aqueous extravaganza coincides with his retrospective, Eliasson's dominance of New York's art scene will be total.

**Christopher Turner** is a writer based in New York.



Olafur Eliasson's *360° Room For All Colours* explores how our minds perceive the external world.

PRIVATE COLLECTION/TANYA BONAKDAR GALLERY, NEW YORK; © 2008 O. ELIASSON



## ESSAY

# The evolution of music

In the second of a nine-part essay series, **Josh McDermott** explores the origins of the human urge to make and hear music.

We think we understand why we are driven to eat, drink, have sex, talk and so forth, based on the uncontroversial adaptive functions of these urges. The drive to engage in music, a compulsion that is arguably just as pervasive in our species, has no such ready explanation. Music was one human behaviour that Charles Darwin was uncertain he could explain, writing in *The Descent of Man, and Selection in Relation to Sex*: "As neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man ... they must be ranked amongst the most mysterious with which he is endowed."

Music's origins have remained puzzling in the years since, although there is no shortage of speculation on the subject. Some argue that music is merely a side effect of traits that evolved for other functions. Our perceptual and cognitive abilities may have accidentally resulted in a system that finds pleasure and interest in musical stimulation. This idea should perhaps be the null hypothesis, and is by no means implausible. Music's perceptual basis could derive from general-purpose auditory mechanisms, its syntactic components could be co-opted from language, and its effect on our emotions could be driven by the acoustic similarity of music to other sounds of greater biological relevance, such as speech or animal vocalizations.

## Just-So stories

Others contend that music is an adaptation with an associated biological function. Some (including Darwin) have proposed that music is the product of sexual selection. Alternative hypotheses include that music promoted social cohesion in group activities such as war or religion, that music was an evolutionary antecedent to language, or that its evolution was driven by the pacifying effect it has on infants.

These discussions run the risk of being mostly 'Just-So' stories, as there are few data with which to test or constrain theories. The

fossil record provides scant clues, and many of the experiments that one might conduct to ask questions about adaptations in other species, such as removal of a trait to test the effect on fitness, are impractical or unethical. Moreover, any current effect on fitness, or lack thereof,

are shared with other cognitive domains. Despite not explicitly testing particular adaptationist theories, these two research directions nonetheless constrain theories of music's evolution in useful ways. Any music-related behaviour that has been selected for must have a genetic basis; and many aspects of music may not be so determined. A trait selected for its role in music might additionally be expected to function specifically in music, and not

in other domains. If some music-related trait — say, the ability to perceive pitch changes — were shown to overlap functionally with something that had a clear adaptive function, such as the perception of speech intonation, it would seem more likely that its role in music is a fortuitous side effect. No single such example is definitive, but repeated demonstrations of nonspecificity would render adaptationist hypotheses less plausible.

## Culture club

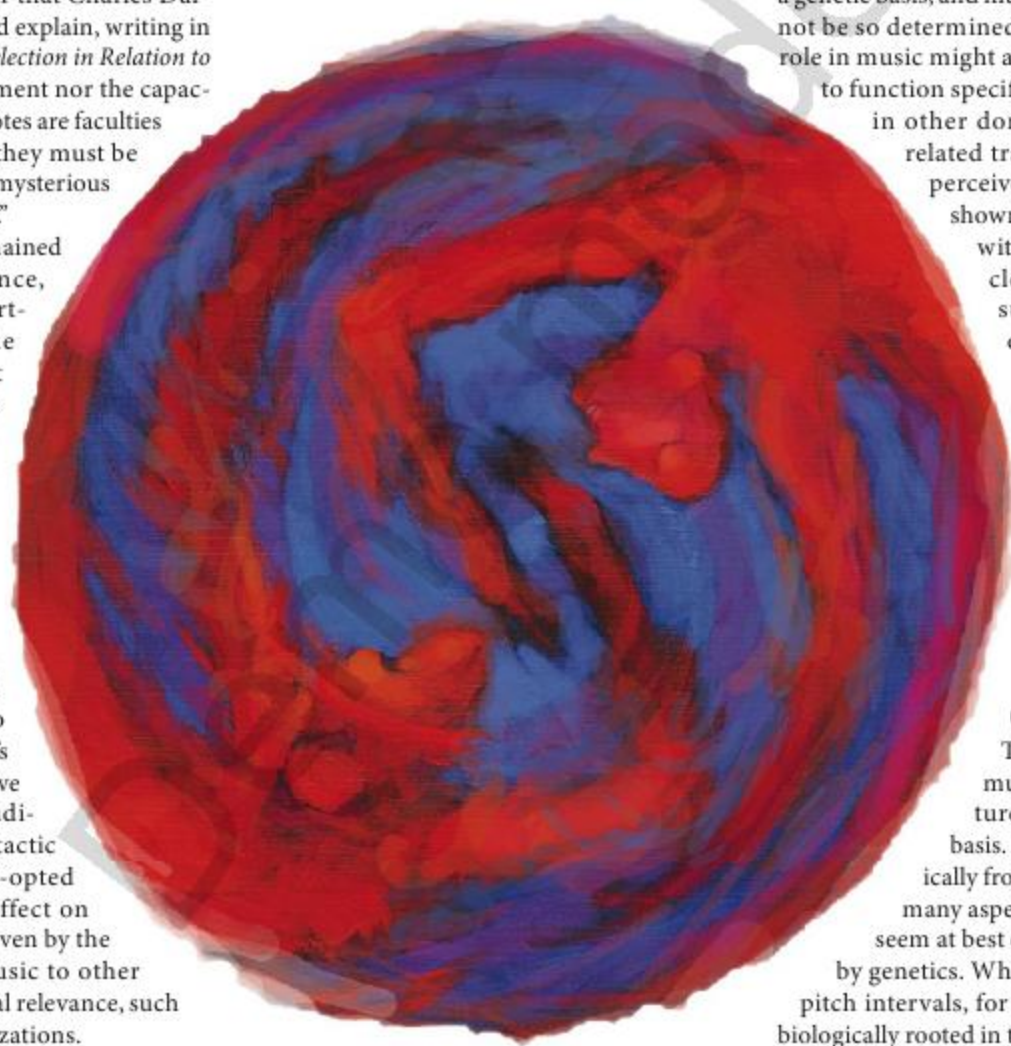
The mere presence of music in every known culture implies some genetic basis. But music varies dramatically from culture to culture, and many aspects of musical behaviour seem at best only weakly constrained by genetics. Whereas our ability to hear pitch intervals, for instance, could well be biologically rooted in the hardware of the auditory system, our emotional response to particular scales or chords seems likely to be acquired from exposure to a particular culture. Interactions between genes and environment are complex, and unravelling their contributions is not easy, but studies of music in different cultures and of musical development offer some hope.

A number of interesting music-related traits emerge in human infants with fairly minimal musical input, providing some evidence for innate constraints. Babies notice when the notes of a

is of questionable relevance, as music's functions today might well differ from those that were operative when it evolved. But just as it is premature to speculate about adaptive function in the absence of data, it is also too soon to conclude that music's origins are fully unknowable.

The way forward lies in targeting two more modest goals. Empirical studies can help to identify musical traits that have a genetic basis, and can test whether these traits

**"Some have proposed that music is the product of sexual selection."**



D. PARKINS



melody are reordered, but not when they are shifted to a different pitch range. Infants, like adults, are sensitive to the relationships between notes, which is preserved in transposition, but altered by reordering. Infants also tend to be captivated by music relative to many other stimuli. Not all music is equivalent to them — they prefer combinations of notes that are judged by adults to sound pleasing, or consonant (the perfect fifth, for instance), over combinations that are less pleasing, or dissonant (a minor second). Infants may even extract metre from music: they react when the rhythm changes from a march to a waltz.

### Universal appeal

Features of music that occur repeatedly around the world despite the substantial cultural variation in music also provide clues to genetically constrained mechanisms. Lullabies seem to qualify as a rare universal — nearly every culture has a genre of music geared towards infants, and there is considerable consistency in how they sound, generally being slow, repetitive and featuring descending pitch contours. Other features that are common, if not completely universal, among cultures include the inclination to dance to music, musical metre, and the hierarchical organization of pitch, giving structural prominence to particular notes over others.

Might any of these potentially heritable traits have evolved specifically for music? Examining functional overlap with other domains can provide insight. Language is generally the most popular candidate; both faculties combine discrete elements (notes/phonemes) into complex structures according to rules. There is growing evidence that this resemblance is more than superficial. Neuroimaging shows frequent overlap between the brain regions that language and music activate. Recent studies suggest that Broca's area — thought to be responsible for linguistic syntax — is also activated by chords that are inappropriate to their musical context, indicating that the area may have a role in musical 'grammar'. Behavioural research also suggests that music and language comprehension can interfere with each other.

It is also conceivable, although largely untested, that much of the interest that young infants have in music is due to its similarity to

speech or other biologically relevant signals. Experiments manipulating this similarity could help to reveal whether there is an effect that is specific to music.

### Animals and music

Music perception in animals can also help to determine the functional specificity of musical traits. Animals lack music; although some birds, whales and other species are said to 'sing', song is absent in most primates, including the apes, and as such is almost surely not homologous to human music. Any music-related trait found in animals, such as

a preference for some sounds over others, or a tendency to hear pitches an octave apart as similar, is thus likely to represent a more general-purpose mechanism. Similarities between human and nonhuman animals can thus indicate that a trait did not evolve for music.

Most studies on music perception in animals

have emphasized their differences with humans rather than potential homologies. Animals, unlike adult and infant humans, have difficulty recognizing melodies that have been shifted up or down in pitch, for instance.

Comparative data are generally more powerful when similarities are present across species rather than differences and, so far, few are evident for music. Many musical capacities, however, remain unexplored. Rhythm and metre, for instance, have been neglected, and could plausibly result from general-purpose mechanisms that some animals might share. The list of species that have been tested

is also short; notably absent are our closest relatives, the apes, whose auditory systems and cognitive abilities are likely to be closest to our own.

### Enduring puzzle

So far, it seems as if there are several aspects of musicality that emerge early in life with minimal exposure to music, and several that are present across diverse musical cultures. These traits may thus be partially heritable. It remains possible that none of these traits is specific to music. This would argue against the idea that music is an adaptation. However, there is much we still do not know, and future experiments could certainly alter this picture.

Music is universal, a significant feature of every known culture, and yet does not serve an obvious, uncontroversial function. As such it stands in contrast to other universal human behaviours. Speculation about its possible adaptive functions has been popular since the time of Darwin, and shows few signs of resolution. Empirical approaches offer a promising alternative. There is no guarantee that a full account of music's origins will ever emerge; in fact, that seems quite unlikely at present. Nonetheless, the right experiments will reveal a great deal — about the innate core of musical behaviour, the traits that might be unique to music, and the possible origins of those components that are not. All of which promises to enrich our appreciation of this human obsession. ■

**Josh McDermott** is a postdoctoral associate studying hearing and music perception in the Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455, USA.

**"Music is a significant feature of every known culture, despite not serving an obvious adaptive function."**



## ESSAY

## 25 years of HIV

Reflecting on how far we have come scientifically since isolating HIV in 1983, **Anthony S. Fauci** urges a renewed commitment to the far greater challenges ahead, especially that of vaccine development.

**T**he HIV/AIDS catastrophe has been one of the defining features of the past quarter of a century. Although it is short-lived in the scheme of public-health crises, the pandemic ranks among the most devastating microbial scourges in human history, one whose full impact has yet to be realized.

Sixty million people have been infected with the human immunodeficiency virus (HIV); nearly half have died, and the toll on families, communities and even entire nations has been profound. Meanwhile, the biomedical research effort directed at HIV/AIDS has resulted in some breathtaking successes. Unlike many other diseases that affect mostly the poor, marginalized and disenfranchised, HIV/AIDS captured the attention of world leaders, the medical, public-health and activist communities, funding agencies, philanthropists and many celebrities. This resulted in an unprecedented scientific and public-health response to the disease, and in welcome attention to some of the many other problems endemic in those populations most severely afflicted with HIV/AIDS, such as malaria, tuberculosis and gender inequality.

Much remains to be accomplished in the global fight against HIV. There are many more scientific and medical hurdles to be cleared and numerous logistical and operational obstacles to making therapies and other interventions available to poor countries, where per capita income is sometimes only a few hundred dollars a year and health-care spending a tiny fraction of that. Reflecting on the era of HIV/AIDS, we must learn from our mis-steps, build on our successes in treatment and prevention, and renew our commitment to developing the truly transforming tools that will one day put this scourge behind us.

### Baffling beginnings

People living through historic events often fail to recognize the significance of what they experience. Such was the case for me, and many of my colleagues, in the first months

#### Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS)

**Abstract.** A retrovirus belonging to the family of recently discovered human T-cell leukemia viruses (HTLV), but clearly distinct from each previous isolate, has been isolated from a Caucasian patient with signs and symptoms that often precede the acquired immune deficiency syndrome (AIDS). This virus is a typical type-C RNA tumour virus, buds from the cell membrane, prefers magnesium for reverse transcription activity, and has an internal antigen epitope similar to HTLV p24. Antibodies from serum of this patient react with proteins from viruses of the new isolate. The type-specific antisera to HTLV-I do not precipitate proteins of the new isolate. The virus from this patient has been transmitted into cord blood lymphocytes, and the virus produced by these cells is similar to the original isolate. From these studies it is concluded that this virus as well as the previous HTLV isolates belong to a general family of T-lymphotropic retroviruses that are horizontally transmitted in humans and may be involved in several pathological syndromes, including AIDS.

The acquired immune deficiency syndrome (AIDS) has recently been recognized in several countries (1). The disease has been reported mainly in homosexual males with multiple partners, and epidemiological studies suggest horizontal transmission by sexual routes (2) as well as by intravenous drug administration (3), and blood transfusion (4). The pronounced depression of cellularity that occurs in AIDS and the

of the AIDS pandemic.

Only in retrospect can we identify its different stages. The first began in June 1981, when physicians in New York and California reported unusual clusters of rare diseases in previously healthy gay men, notably *Pneumocystis carinii* pneumonia and a form of cancer called Kaposi's sarcoma. When we in the medical profession read those initial cases 27 years ago — and treated some of the early AIDS patients — our prevailing emotion was bewilderment.

I had seen other 'mystery' diseases in my career, such as the legionnaires' disease outbreak of 1976. But AIDS was from the beginning much more insidious and enigmatic. As cases began to appear among distinctly different social groups in 1981–82, and as we began to understand better the profound and complex immunodeficiency of our patients, it became clear that we were witnessing the unfolding of something truly novel and frightening. The severity of AIDS and the signs that it apparently could be spread by a ubiquitous human activity — sex — suggested that we were in for a difficult time.

The scientific community marshalled its resources and talent to fight AIDS; investiga-

tors from different disciplines began working on this new disease. Within months of the recognition of the first cases in the summer of 1981, I shifted the direction of my laboratory from the study of inflammatory diseases to this curious new syndrome. My research has been closely intertwined with HIV and AIDS ever since. For nearly two years, the cause of AIDS remained elusive; the scientific community was largely baffled, lacking good leads for developing therapies or even a diagnostic test.

Those of us caring for patients with AIDS had few tools at our disposal. The only treatments we could provide were largely palliative and the lifespan of most of our patients was measured in months. Those years were the darkest of my professional career.

### Glimmer of hope

Twenty-five years ago this month came a glimmer of hope. In 1983, Luc Montagnier's research team in Paris published in *Science* the first paper (pictured) providing evidence linking a retrovirus to AIDS. The following year, further data from Robert Gallo's group in the United States provided convincing evidence that this retrovirus (later named HIV) was the cause of AIDS. That these two outstanding scientists became embroiled in a controversy — largely played out in the media — about who discovered HIV was an unfortunate distraction. As they would later write in *The New England Journal of Medicine*: "Many lessons can be drawn from this early intense period, and most suggest that science requires greater modesty." A quarter of a century on, the importance of collaboration, collegiality and, yes, modesty, are ever more apparent, as it becomes clear that no single research group or discipline will solve the puzzles of HIV/AIDS.

After the discovery of HIV, research moved at a breathtaking pace. A blood test to diagnose patients and to screen the blood supply quickly followed, as did enormous progress in understanding the genetics and structure of HIV and its disease-causing mechanisms. The rapid clinical testing and licensing in 1987 of the first effective drug against HIV, zidovudine (AZT), caused great excitement. In retrospect this was unfounded, as the molecular characteristics

**"HIV/AIDS is predominantly a disease of the developing world, where access to scientific advances and therapies is difficult."**



of HIV, notably its propensity to replicate and mutate rapidly, made any single drug unlikely to hold the virus in check. Previous experience with antimicrobials for other diseases and the inevitable emergence of drug-resistant pathogens should have made us more cautious about the prospects for AZT monotherapy.

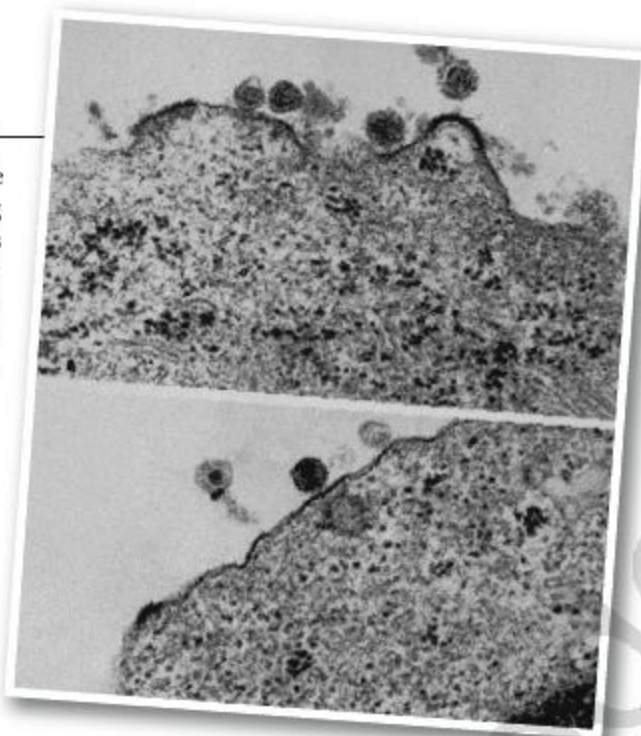
HIV quickly developed resistance to AZT and the benefits of the drug rapidly waned. Initial optimism about therapy gave way to sobering reality as the AIDS pandemic continued to grow in the United States and elsewhere. Clinicians remained hobbled by a lack of effective anti-HIV drug regimes, and many more patients were lost to AIDS.

Gradually, the fruits of cutting-edge drug development began to appear. In late 1995, the first of a new class of antiretroviral drugs — protease inhibitors — reached the market. Other new drugs that attacked the virus in different ways followed, and we soon had a greater number of effective drugs for HIV than for all other viral diseases combined. The new therapies used in combination with older medicines rapidly improved the prognosis for vast numbers of HIV-infected patients. The AIDS death rate in the United States fell by more than two-thirds within two years of the licensing of the first protease inhibitor. Despite certain limitations of the new treatments, notably toxicity and drug resistance, they launched a new era of optimism.

But HIV/AIDS is predominantly a disease of the developing world, where access to scientific advances and therapies is difficult. Fewer than one-third of the people who need antiretroviral therapy are currently receiving it, despite heroic efforts on the part of individuals and organizations, and some truly transforming and innovative programmes such as the Global Fund to Fight AIDS, Tuberculosis, and Malaria and the US President's Emergency Plan for AIDS Relief. Furthermore, it is clear that treatment alone will never end the AIDS pandemic. New infections far outstrip our ability to treat everyone infected with the virus: around three people are newly infected for every person put on therapy — and current HIV therapy is a lifelong commitment.

### More mountains

To improve these formidable odds, we have two main options. The first is to cure patients, that is, to purge every vestige of virus from their bodies so that a course of treatment could be measured in weeks or months rather than



**Agent identified:** early images captured by Robert Gallo (top) and Luc Montagnier of the retrovirus later named HIV.

a lifetime. Sadly, because of the ability of HIV to hide within cells from both drugs and the immune system, such a treatment regime has proved elusive, although important work in this area is being pursued.

This leaves us with the second option: preventing HIV infection in the first place. We have a proven array of HIV-prevention and harm-reduction strategies: behavioural modification; condom distribution; antiretroviral drug regimes to prevent HIV transmission from mother to baby; and the provision of clean needles and syringes to drug users. We must now do better at delivering prevention: less than 20% of those at risk of HIV infection are currently receiving such help.

Encouragingly, new means of preventing HIV are emerging. Large randomized and controlled clinical trials in Africa suggest that adult male circumcision, if properly and hygienically performed and accompanied by appropriate counselling

and post-surgical care, can help prevent men becoming infected with HIV by heterosexual intercourse. Ongoing work to develop microbicide gels or creams to be applied before sex offers the hope of empowering women to protect themselves from HIV infection when the use of condoms or the refusal of sexual intercourse is not feasible.

Yet the major goal of HIV/AIDS research eludes us: the development of a safe and effective HIV vaccine, our best hope for ultimately ending the pandemic. The search for a vaccine has been made extremely difficult by the nature

of the virus, particularly its ability to integrate into the genome of host cells, to mutate readily and to conceal that part of its outer coat that would induce protective antibodies.

It is now clear that we were naive to think there would be a straight path from the discovery and characterization of HIV to the development of a vaccine. HIV has proved very different from those viruses for which we have developed effective immunizations. We must solve the mystery of how to prompt the human body to produce a protective immune response that is even better than the one elicited by natural infection. This will require a commitment to fundamental research to address the many questions that remain about HIV and its interactions with its human host. HIV/AIDS science, particularly that involving a vaccine, is in some ways still in its infancy. We must move forward by fostering creative thinking over many different disciplines.

In this regard, the pursuit of new research avenues by established scientists and especially by younger investigators is critical. In addition to the disciplines classically associated with HIV research — virology and immunology — we must encourage more 'cross-fertilization' with other fields such as genetics, structural biology, systems biology and peptide chemistry as we strive to generate the knowledge needed to develop an HIV vaccine.

Delivering HIV interventions for the people it most affects requires political will, a long-term supply of considerable financial resources, scientific and public-health vision, and dedication from all sectors of society. With these ingredients, the trajectory of our fight against the HIV/AIDS pandemic in the next quarter of a century could move from cautious optimism towards triumph. Absent any of these factors, and history will not judge us kindly. ■

**Anthony S. Fauci** is director of the National Institute of Allergy and Infectious Diseases (NIAID) at the US National Institutes of Health in Bethesda, Maryland, and chief of the NIAID Laboratory of Immunoregulation.

### FURTHER READING

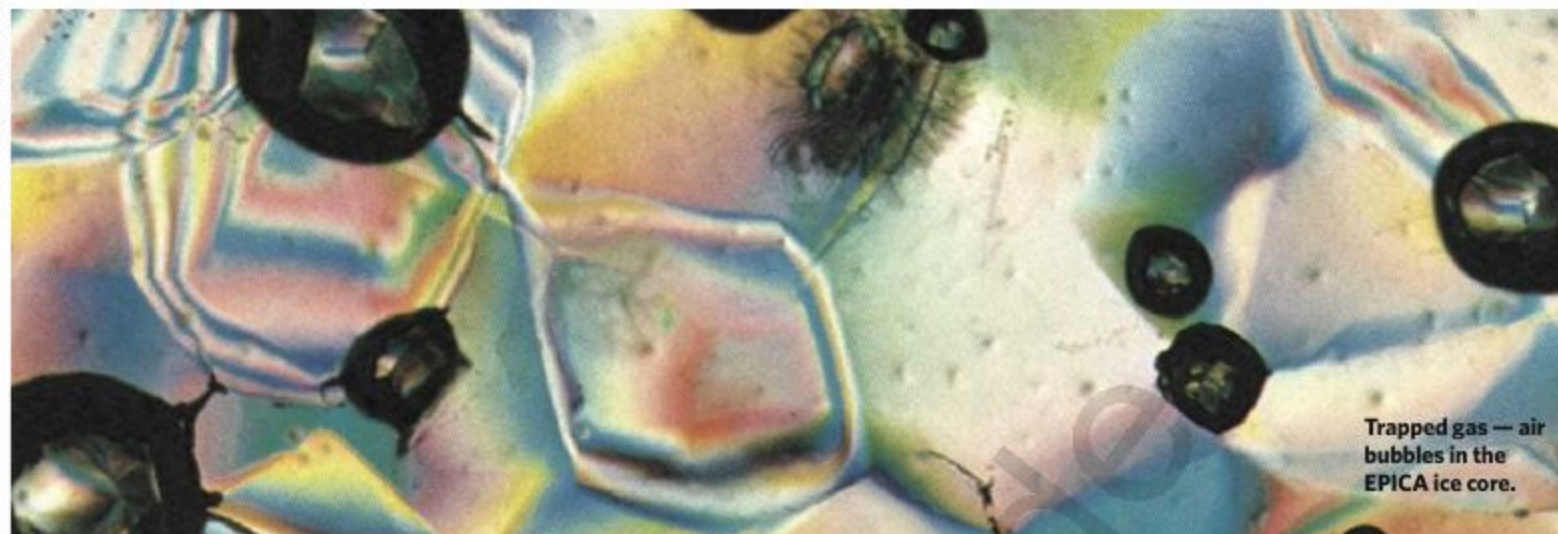
- Barré-Sinoussi, F. *et al.* *Science* **220**, 868–871 (1983).
- Fauci, A. S. & Lane, H. C. in *Harrison's Principles of Internal Medicine*, 17th edition (eds Fauci, A. S. *et al.*) 1137–1204 (McGraw Hill, New York, 2008).
- Fauci, A. S. *Science* **239**, 617–622 (1988).
- Gallo, R. C. & Montagnier, L. N. *Engl. J. Med.* **349**, 2283–2285 (2003).
- Joint United Nations Programme on HIV/AIDS (UNAIDS) *AIDS Epidemic Update: December 2007* (UNAIDS, Geneva, 2007).
- Levy, J. A. *HIV and the Pathogenesis of AIDS* (ASM Press, Washington DC, 2007).

**"The major goal of HIV/AIDS research eludes us: the development of a safe and effective HIV vaccine."**



## NEWS &amp; VIEWS

THE EPICA COLLABORATION



Trapped gas — air bubbles in the EPICA ice core.

## PALAEOCLIMATE

# Windows on the greenhouse

Ed Brook

**Data laboriously extracted from an Antarctic ice core provide an unprecedented view of temperature, and levels of atmospheric carbon dioxide and methane, over the past 800,000 years of Earth's history.**

Palaeoclimatologists are scientific detectives. Using indirect clues from concentrations of stable isotopes and trace elements, and from fossils and other components of the geological record, they infer changes in climate long before they themselves were on the scene. Direct evidence of past environmental conditions is rare, which makes it all the more valuable where it does occur. In this issue<sup>1,2</sup>, members of the EPICA (European Project for Ice Coring in Antarctica) collaboration present the latest, and longest, record from perhaps the most valuable of these archives: the atmospheric gases trapped and preserved in ice cores extracted from Earth's polar regions.

Polar ice cores provide us with the long view of the cycling of greenhouse gases such as carbon dioxide and methane. Their potential is being realized by a relatively small band of international scientists who are gradually drilling further down into the ice cap and progressively analysing older ice cores. Until recently, the Vostok ice core from eastern Antarctica set the benchmark<sup>3</sup> — an iconic 440,000-year data set that became a central backdrop for discussions about modern climate change.

That ante was upped in 2005 by a 650,000-year record<sup>4,5</sup> from EPICA's 'Dome C', another drilling site in eastern Antarctica where much older ice could be extracted. An 800,000-year reconstruction of temperature change from the core already existed<sup>6</sup>. Now, after years of careful work and collaboration, Dome C

has yielded a complete reconstruction of the history of atmospheric carbon dioxide (Lüthi *et al.*, page 379)<sup>1</sup> and methane (Louergue *et al.*, page 383)<sup>2</sup> over the past 800,000 years.

The fundamental conclusion that today's concentrations of these greenhouse gases have no past analogue in the ice-core record remains firm. The general long-term behaviour of methane and carbon dioxide, following patterns driven ultimately by slow changes in Earth's orbit, continues throughout the older sections of the records. The remarkably strong correlations of methane and carbon dioxide with temperature reconstructions also stand (Fig. 1, overleaf).

The data further reinforce the tight link between greenhouse gases and climate, a link maintained by as-yet only partially understood feedbacks in the Earth system. Variations in methane levels are most probably caused by variations in the influence of temperature and rainfall on wetlands in the tropics and boreal (high-northern-latitude) regions. Carbon dioxide variability is almost universally viewed as an oceanic phenomenon, a consequence of the large pools of carbon sequestered there. Changes in ocean circulation, biological productivity, carbon dioxide solubility and other aspects of ocean chemistry have been implicated, but the exact mix of mechanisms is not clear.

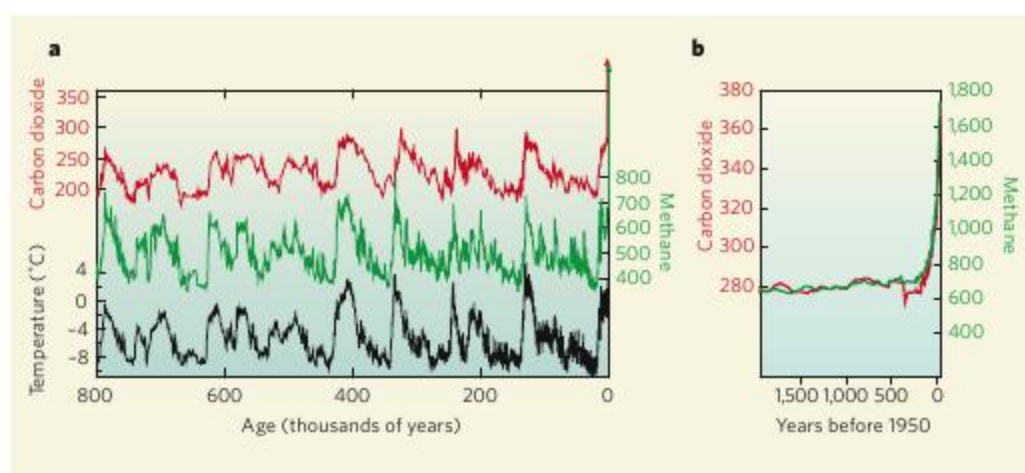
In considering these extended records<sup>1,2</sup> in detail, intriguing nuances emerge. Their most prominent feature is a sawtooth-shaped

variability on 100,000-year timescales. As reported last year<sup>6</sup>, the amplitude of the 100,000-year temperature cycle reconstructed at Dome C seems to have changed fundamentally about 450,000 years ago (Fig. 1). Warm phases (interglacials) in the later period have been warmer, whereas cold phases (glacials) seem similar throughout the record. The carbon dioxide record generally shares this pattern, with muted cycles in its older part. Methane also follows the trend, though not as strongly: relatively high methane maxima in the oldest interglacial cycle approach those of later warm periods.

A curious facet of the extended carbon dioxide record is unusually low levels of the gas during the two earliest glacial-interglacial cycles. Lüthi *et al.* speculate<sup>1</sup> that, taken as a whole, the carbon dioxide record is hinting at a longer-term cycle in mean levels of the gas that takes 400,000–500,000 years to complete. The eccentricity of Earth's orbit — its deviation from a perfect circle — does vary with a 413,000-year period. But whether this or some other mechanism explains any variation awaits the retrieval of an even older ice core.

The extended records also provide information about shorter-term, millennial-scale climate change taking place within the longer cycles. Data from ice cores in Greenland covering the past 110,000 years show that variations in methane levels were extremely closely coupled to episodes of abrupt warming and cooling in the mid-to-high latitudes of the Northern





**Figure 1 | A long look back.** **a**, The 800,000-year records of atmospheric carbon dioxide (red; parts per million, p.p.m.) and methane (green; parts per billion, p.p.b.) from the EPICA Dome C ice core<sup>1,2</sup>, together with a temperature reconstruction (relative to the average of the past millennium) based on the deuterium–hydrogen ratio of the ice<sup>6</sup>, reinforce the tight coupling between greenhouse-gas concentrations and climate observed in previous, shorter records. The 100,000-year ‘sawtooth’ variability undergoes a change about 450,000 years ago, with the amplitude of variation, especially in the carbon dioxide and temperature records, greater since that point than it was before. Concentrations of greenhouse gases in the modern atmosphere are highly anomalous with respect to natural greenhouse-gas variations (present-day concentrations are around 380 p.p.m. for carbon dioxide and 1,800 p.p.b. for methane). **b**, The carbon dioxide and methane trends from the past 2,000 years<sup>13,14</sup>.

Hemisphere<sup>7,8</sup>. No older records from Greenland exist at present; indeed, records extending back further than about 200,000 years are not expected to be found there owing to high accumulation rates and the flow of older ice towards the margins of the ice sheet. But assuming that the close coupling between Greenland’s temperature and levels of atmospheric methane holds before 110,000 years ago, jumps in the Dome C methane record provide a Southern Hemisphere proxy for abrupt warming in the Northern Hemisphere. Loulergue *et al.*<sup>2</sup> identify 74 such jumps in their data and, following this logic, conclude that abrupt warming and cooling in Greenland and the Northern Hemisphere has been a characteristic of the climate system over at least the past 800,000 years.

Again using methane as a proxy for Greenland’s temperature patterns, it can be shown that on millennial timescales carbon dioxide concentrations rose during times when Greenland was cold. At the same time, Antarctica warmed<sup>9</sup>. This pattern has been attributed to the effect of changes in ocean circulation on the carbon cycle and climate<sup>10</sup>. Lüthi *et al.*<sup>1</sup> identify examples of this kind of variability in ice as old as 750,000 to 780,000 years, another indication that these millennial patterns pervade the palaeoclimate record.

These new benchmark data<sup>1,2</sup> for greenhouse-gas variability pose questions as to what a much longer record might show. One such question is whether the 400,000–500,000-year cycle speculated on by Lüthi *et al.*<sup>1</sup> is a real effect. Another is whether the 100,000-year cycles in carbon dioxide and methane, now so clearly established, give way to 40,000-year cycles before about 900,000 years ago; such behaviour might be predicted by comparison with climate reconstructions from ocean sediments<sup>11</sup>. If that is the case, what caused

the shift? Was it a reduction in mean concentrations of greenhouse gases 900,000 years ago? This commonly cited theory<sup>12</sup> is just one of many competing hypotheses<sup>11</sup>.

The international community of ice-core scientists, under the auspices of the umbrella

group IPICS (International Partners in Ice Core Sciences), has set itself the immediate target of establishing a continuous 1.5-million-year record to attempt to answer these questions. The search for the right sites is beginning, and is likely to take several years. The best places are undoubtedly in eastern Antarctica, most probably in remote, high regions where snowfall rates and temperatures are extremely low. Meeting the challenge of drilling those cores should open up a further window on goings-on in the greenhouse.

Ed Brook is in the Department of Geosciences, Oregon State University, 104 Wilkinson Hall, Corvallis, Oregon 97331-5506, USA.

e-mail: brooke@science.oregonstate.edu

1. Lüthi, D. *et al.* *Nature* **453**, 379–382 (2008).
2. Loulergue, L. *et al.* *Nature* **453**, 383–386 (2008).
3. Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
4. Spahni, R. *et al.* *Science* **310**, 1317–1321 (2005).
5. Siegenthaler, U. *et al.* *Science* **310**, 1313–1317 (2005).
6. Jouzel, J. *et al.* *Science* **317**, 793–796 (2007).
7. Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. *Nature* **391**, 141–146 (1998).
8. Huber, C. *Earth Planet. Sci. Lett.* **243**, 504–519 (2006).
9. Ahn, J. & Brook, E. *J. Geophys. Res. Lett.* **34**, L10703 (2007).
10. Schmittner, A., Brook, E. J. & Ahn, J. in *Ocean Circulation: Mechanisms and Impacts* (eds Schmittner, A., Chiang, J. C. H. & Hemming, S.) 315–334 (AGU, Washington DC, 2007).
11. Clark, P. U. *et al.* *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
12. Berger, A., Li, X. S. & Loutre, M.-F. *Quat. Sci. Rev.* **18**, 1–11 (1999).
13. MacFarling Meure, C. *et al.* *Geophys. Res. Lett.* **33**, L14810 (2006).
14. [www.cmdl.noaa.gov/infodata/ftpdata.html](http://www.cmdl.noaa.gov/infodata/ftpdata.html)

## SIGNAL TRANSDUCTION

# The rhodopsin story continued

Gebhard F. X. Schertler

**Determination of the architecture of an invertebrate photoreceptor protein, squid rhodopsin, is a notable event. It illuminates the mechanism of invertebrate vision and a ubiquitous intracellular signalling system.**

Many invertebrates have excellent visual systems<sup>1</sup>. Squid, for example, are formidable hunters that rely on their acute visual abilities to catch their prey. As in vertebrates, the properties of the photoreceptor protein rhodopsin contribute significantly to those abilities. So one reason for the attention that will be paid to the paper on page 363 of this issue<sup>2</sup>, in which Murakami and Kouyama present a high-resolution crystal structure of rhodopsin from the squid retina, is that it will help in understanding invertebrate vision. But the paper’s significance extends far beyond that.

Rhodopsin is located in the cell membrane of photoreceptor cells. When activated by light, it undergoes a conformational change that triggers the action of a heterotrimeric GTP-binding protein (G protein) lying just beneath the cell membrane. Rhodopsin, therefore, is a member of the superfamily of G-protein-coupled receptors (GPCRs), all of which contain

seven structural domains that each span the membrane. However, invertebrate rhodopsin signals through the  $\alpha$ -subunit of a  $G_q$  type of G protein (rather than transducin, the  $\alpha$ -subunit for vertebrate rhodopsin), leading to activation of phospholipase C and eventually the opening of a calcium channel (rather than activation of cyclic GMP phosphodiesterase, leading eventually to closing of a cation channel, which occurs in vertebrates<sup>3</sup>).

In other words, Murakami and Kouyama<sup>2</sup> are the first to determine the structure of a  $G_q$ -coupled GPCR. The wider significance of the paper is that many hormone and neurotransmitter receptors signal through  $G_q$ , including vasopressin and oxytocin receptors as well as serotonin and acetylcholine receptors in the brain. And many drugs, such as anti-histamines and angiotensin antagonists, target  $G_q$ -coupled receptors.

The functional unit of squid rhodopsin

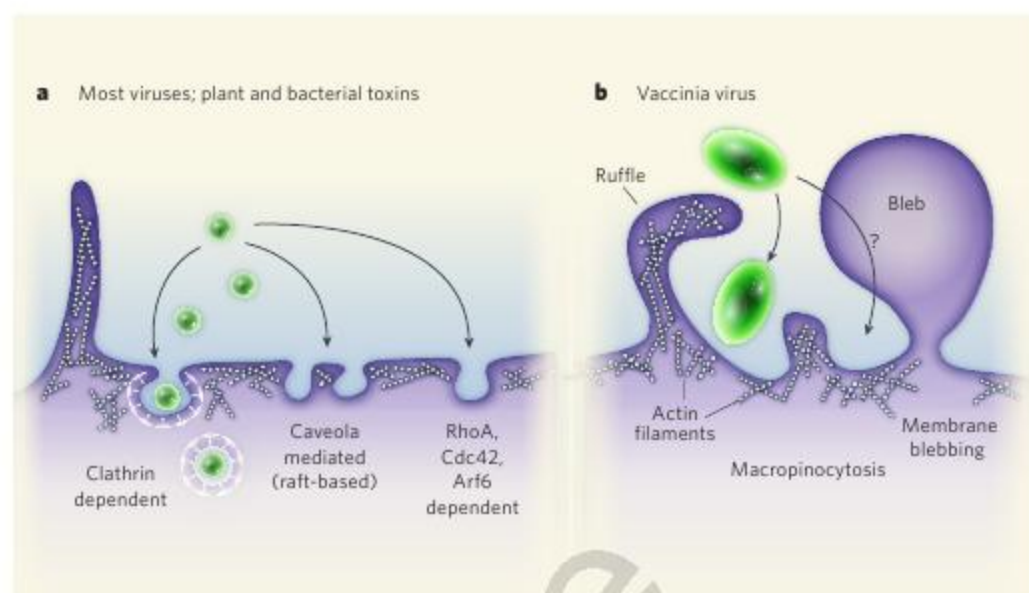


underlying mechanism is as follows. When a cell becomes infected with the virus, it displays apoptotic features, including the presence of phosphatidylserine in its outer membrane layer. Thus, when the virus buds off from the cell, it inherits this as part of its envelope. Consequently, cells probably 'mistake' the unusually large vaccinia virus for an apoptotic body (the debris of dying cells) and engulf it.

Mercer and Helenius<sup>3</sup> find that vaccinia virus seems to enter its host cell through an endocytic process called macropinocytosis, which normally mediates fluid uptake. Like virus budding, virus uptake also exploits apoptotic mechanisms. The authors show that vaccinia virus initially binds to cytoplasmic protrusions called filopodia that extend from the surface of the target cell. It moves along them towards the cell body, and then somehow sends signals into the cell, stimulating extensive membrane blebbing. During blebbing, the actin network that forms the scaffolding of the cell beneath the cell membrane becomes detached (Fig. 2). Blebbing is transient, and soon after, the network reassembles in the same cellular location, and the blebs retract<sup>5</sup>.

The authors find that virus internalization by this macropinocytosis-like process is mediated by the blebs, as bleb retraction and re-formation of the actin network coincide with virus entry. Moreover, the drug blebbistatin, which inhibits blebbing<sup>6</sup>, blocks virus entry. In addition, virus internalization requires several proteins (including actin, PAK1, Rac1 and various lipid- and protein-kinase enzymes) that are involved in membrane blebbing<sup>3</sup>. Thus, blebbing might participate in endocytosis, probably when the bleb is retracting and the actin system is re-forming, as the bleb could fold over, or invaginate — a process that would resemble macropinocytosis.

The possibility that blebbing and macropinocytosis are two entirely independent



**Figure 2 | Vaccinia virus chooses to be different.** **a**, Most viruses, as well as plant and bacterial toxins, enter host cells through the classic method of endocytosis, which involves membrane invagination and pinching off of the membrane to form an intracellular transport vesicle. Three different forms of this type of endocytosis are shown. **b**, Mercer and Helenius<sup>3</sup> find that vaccinia virus enters by an endocytic mechanism resembling macropinocytosis. On the cell surface, the virus triggers membrane blebbing, which might also lead to the formation of membrane invaginations that will evolve into transport vesicles. The actin network, which is normally present beneath the cell membrane and is involved in various endocytic processes, is absent in the bleb, but re-forms during bleb retraction.

processes, both of which are stimulated by vaccinia virus, is equally valid. Specifically, blebbistatin inhibits the myosin II protein, which is required for blebbing. But it can also inhibit myosin-II-independent processes such as macropinocytosis<sup>7</sup>. So Mercer and Helenius's observations raise the question of whether macropinocytosis should be subdivided into at least two types: the traditional type in which membrane ruffling (small, dynamic folds of the membrane; Fig. 2b) precedes vesicle formation; and the type that involves blebbing. Multi-modal macropinocytosis would fit well with the increasing number of other types of endocytosis that are being identified<sup>8</sup>.

That opportunistic pathogens exploit various mechanisms for entry and replication within host cells is also documented in a study<sup>9</sup> of the bacterium *Pseudomonas aeruginosa*. This pathogen induces the formation of very large membrane blebs in epithelial cells, entering the blebs and replicating there. The blebs are quite translucent, and do not seem to contain cytoskeletal elements such as actin. Moreover, the bacteria are highly motile within the blebs. But the exact mechanism of bacterial entry into them remains elusive.

Discoveries often raise new questions, and Mercer and Helenius's work<sup>3</sup> is no exception. First, what is the exact relationship between blebbing and macropinocytosis? Cholesterol, for example, is required for both virus infection and macropinocytosis. Is it also required for blebbing? Is virus-induced blebbing cell-type-specific? What happens in polarized cells, in which membrane components and structural elements vary in different parts of the cell, as opposed to the non-polarized cell lines that were studied here? As the active form of

the Arf6 protein inhibits virus infection, one might also wonder how Arf6 is involved in this process. Much is to be learnt about the mechanisms and pathways underlying the internalization of opportunistic pathogens such as vaccinia virus. Ironically, further knowledge about endocytosis itself is likely to come from studies of pathogens.

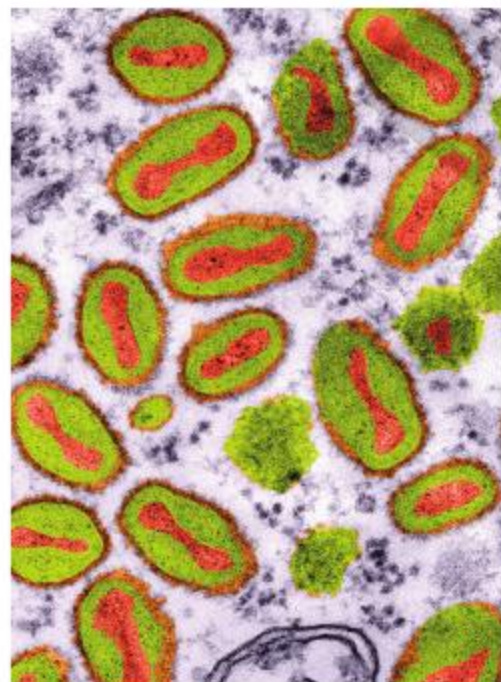
Kirsten Sandvig is in the Centre for Cancer Biomedicine, University of Oslo, and the Department of Biochemistry, Institute for Cancer Research, Norwegian Radium Hospital, Rikshospitalet University Hospital, Montebello, 0310 Oslo, Norway. Bo van Deurs is at the Institute of Cellular and Molecular Medicine, The Panum Building, Faculty of Health Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark.

e-mails: ksandvig@radium.uio.no;  
bvd@sund.ku.dk

1. Marsh, M. & Helenius, A. *Cell* **124**, 729–740 (2006).
2. Sandvig, K. & van Deurs, B. *Gene Ther.* **12**, 865–872 (2005).
3. Mercer, J. & Helenius, A. *Science* **320**, 531–535 (2008).
4. Henson, P. M., Bratton, D. L. & Fadok, V. A. *Curr. Biol.* **11**, R795–R805 (2001).
5. Charras, G. T., Hu, C.-K., Coughlin, M. & Mitchison, T. J. *J. Cell Biol.* **175**, 477–490 (2006).
6. Straight, A. F. et al. *Science* **299**, 1743–1747 (2003).
7. Shu, S., Liu, X. & Korn, E. D. *Proc. Natl Acad. Sci. USA* **102**, 1472–1477 (2005).
8. Sandvig, K., Torgersen, M. L., Raa, H. A. & van Deurs, B. *Histochem. Cell Biol.* **129**, 267–276 (2008).
9. Angus, A. A. et al. *Infect. Immun.* **76**, 1992–2001 (2008).

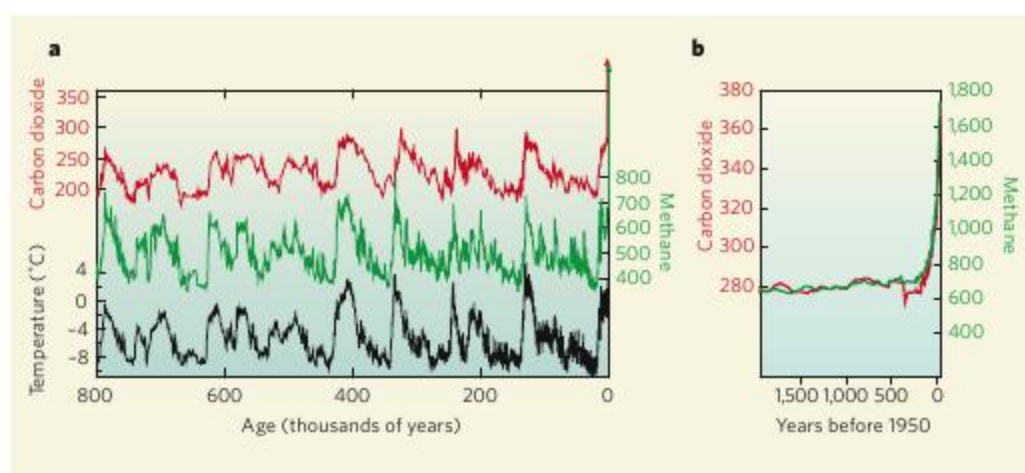
#### Correction

In the News & Views article "Palaeoclimate: Windows on the greenhouse" by Ed Brook (*Nature* **453**, 291–292; 2008), the wrong credit was used for the picture on page 291. The figure in fact came from the 1978 doctoral thesis of W. Berner (University of Bern).



**Figure 1 | Source of infection.** False-colour electron microscopy image of vaccinia virus.





**Figure 1 | A long look back.** **a**, The 800,000-year records of atmospheric carbon dioxide (red; parts per million, p.p.m.) and methane (green; parts per billion, p.p.b.) from the EPICA Dome C ice core<sup>1,2</sup>, together with a temperature reconstruction (relative to the average of the past millennium) based on the deuterium–hydrogen ratio of the ice<sup>6</sup>, reinforce the tight coupling between greenhouse-gas concentrations and climate observed in previous, shorter records. The 100,000-year ‘sawtooth’ variability undergoes a change about 450,000 years ago, with the amplitude of variation, especially in the carbon dioxide and temperature records, greater since that point than it was before. Concentrations of greenhouse gases in the modern atmosphere are highly anomalous with respect to natural greenhouse-gas variations (present-day concentrations are around 380 p.p.m. for carbon dioxide and 1,800 p.p.b. for methane). **b**, The carbon dioxide and methane trends from the past 2,000 years<sup>13,14</sup>.

Hemisphere<sup>7,8</sup>. No older records from Greenland exist at present; indeed, records extending back further than about 200,000 years are not expected to be found there owing to high accumulation rates and the flow of older ice towards the margins of the ice sheet. But assuming that the close coupling between Greenland’s temperature and levels of atmospheric methane holds before 110,000 years ago, jumps in the Dome C methane record provide a Southern Hemisphere proxy for abrupt warming in the Northern Hemisphere. Loulergue *et al.*<sup>2</sup> identify 74 such jumps in their data and, following this logic, conclude that abrupt warming and cooling in Greenland and the Northern Hemisphere has been a characteristic of the climate system over at least the past 800,000 years.

Again using methane as a proxy for Greenland’s temperature patterns, it can be shown that on millennial timescales carbon dioxide concentrations rose during times when Greenland was cold. At the same time, Antarctica warmed<sup>9</sup>. This pattern has been attributed to the effect of changes in ocean circulation on the carbon cycle and climate<sup>10</sup>. Lüthi *et al.*<sup>1</sup> identify examples of this kind of variability in ice as old as 750,000 to 780,000 years, another indication that these millennial patterns pervade the palaeoclimate record.

These new benchmark data<sup>1,2</sup> for greenhouse-gas variability pose questions as to what a much longer record might show. One such question is whether the 400,000–500,000-year cycle speculated on by Lüthi *et al.*<sup>1</sup> is a real effect. Another is whether the 100,000-year cycles in carbon dioxide and methane, now so clearly established, give way to 40,000-year cycles before about 900,000 years ago; such behaviour might be predicted by comparison with climate reconstructions from ocean sediments<sup>11</sup>. If that is the case, what caused

the shift? Was it a reduction in mean concentrations of greenhouse gases 900,000 years ago? This commonly cited theory<sup>12</sup> is just one of many competing hypotheses<sup>11</sup>.

The international community of ice-core scientists, under the auspices of the umbrella

group IPICS (International Partners in Ice Core Sciences), has set itself the immediate target of establishing a continuous 1.5-million-year record to attempt to answer these questions. The search for the right sites is beginning, and is likely to take several years. The best places are undoubtedly in eastern Antarctica, most probably in remote, high regions where snowfall rates and temperatures are extremely low. Meeting the challenge of drilling those cores should open up a further window on goings-on in the greenhouse.

Ed Brook is in the Department of Geosciences, Oregon State University, 104 Wilkinson Hall, Corvallis, Oregon 97331-5506, USA.

e-mail: brooke@science.oregonstate.edu

1. Lüthi, D. *et al.* *Nature* **453**, 379–382 (2008).
2. Loulergue, L. *et al.* *Nature* **453**, 383–386 (2008).
3. Petit, J. R. *et al.* *Nature* **399**, 429–436 (1999).
4. Spahni, R. *et al.* *Science* **310**, 1317–1321 (2005).
5. Siegenthaler, U. *et al.* *Science* **310**, 1313–1317 (2005).
6. Jouzel, J. *et al.* *Science* **317**, 793–796 (2007).
7. Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. *Nature* **391**, 141–146 (1998).
8. Huber, C. *Earth Planet. Sci. Lett.* **243**, 504–519 (2006).
9. Ahn, J. & Brook, E. *J. Geophys. Res. Lett.* **34**, L10703 (2007).
10. Schmittner, A., Brook, E. J. & Ahn, J. in *Ocean Circulation: Mechanisms and Impacts* (eds Schmittner, A., Chiang, J. C. H. & Hemming, S.) 315–334 (AGU, Washington DC, 2007).
11. Clark, P. U. *et al.* *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
12. Berger, A., Li, X. S. & Loutre, M.-F. *Quat. Sci. Rev.* **18**, 1–11 (1999).
13. MacFarling Meure, C. *et al.* *Geophys. Res. Lett.* **33**, L14810 (2006).
14. [www.cmdl.noaa.gov/infodata/ftpdata.html](http://www.cmdl.noaa.gov/infodata/ftpdata.html)

## SIGNAL TRANSDUCTION

# The rhodopsin story continued

Gebhard F. X. Schertler

**Determination of the architecture of an invertebrate photoreceptor protein, squid rhodopsin, is a notable event. It illuminates the mechanism of invertebrate vision and a ubiquitous intracellular signalling system.**

Many invertebrates have excellent visual systems<sup>1</sup>. Squid, for example, are formidable hunters that rely on their acute visual abilities to catch their prey. As in vertebrates, the properties of the photoreceptor protein rhodopsin contribute significantly to those abilities. So one reason for the attention that will be paid to the paper on page 363 of this issue<sup>2</sup>, in which Murakami and Kouyama present a high-resolution crystal structure of rhodopsin from the squid retina, is that it will help in understanding invertebrate vision. But the paper’s significance extends far beyond that.

Rhodopsin is located in the cell membrane of photoreceptor cells. When activated by light, it undergoes a conformational change that triggers the action of a heterotrimeric GTP-binding protein (G protein) lying just beneath the cell membrane. Rhodopsin, therefore, is a member of the superfamily of G-protein-coupled receptors (GPCRs), all of which contain

seven structural domains that each span the membrane. However, invertebrate rhodopsin signals through the  $\alpha$ -subunit of a  $G_q$  type of G protein (rather than transducin, the  $\alpha$ -subunit for vertebrate rhodopsin), leading to activation of phospholipase C and eventually the opening of a calcium channel (rather than activation of cyclic GMP phosphodiesterase, leading eventually to closing of a cation channel, which occurs in vertebrates<sup>3</sup>).

In other words, Murakami and Kouyama<sup>2</sup> are the first to determine the structure of a  $G_q$ -coupled GPCR. The wider significance of the paper is that many hormone and neurotransmitter receptors signal through  $G_q$ , including vasopressin and oxytocin receptors as well as serotonin and acetylcholine receptors in the brain. And many drugs, such as anti-histamines and angiotensin antagonists, target  $G_q$ -coupled receptors.

The functional unit of squid rhodopsin

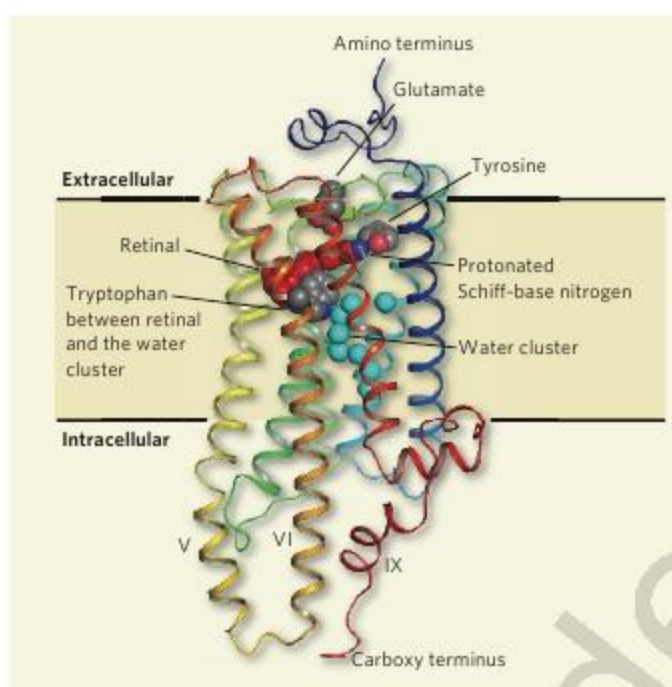


consists of a light-sensitive chromophore, retinal, covalently bound to an evolutionarily conserved lysine amino-acid residue in helix VII of the opsin protein. This covalent bond between a carbon atom from the aldehyde group of retinal and a nitrogen atom from the amino group of the side chain of lysine is called a Schiff base. The protonated nitrogen atom of this bond is thought to require a negatively charged 'counterion' for it to be neutralized.

Activation of rhodopsin works in the following way. Retinal has a double bond in its carbon chain that is isomerized upon absorption of light. This light-induced isomerization of the carbon chain from a kinked to a straightened form leads to a conformational change of the entire protein, triggering exchange of GDP for GTP in the  $G_q$   $\alpha$ -subunit. The GTP-bound form then activates phospholipase C, leading to the formation of inositol phosphates, membrane-bound diacylglycerol and polyunsaturated fatty acids<sup>3</sup>. Further events, which are not completely understood, lead to the opening of a calcium channel and the generation of nerve impulses that carry the relevant visual information to the brain.

Invertebrate rhodopsins are larger than vertebrate rhodopsins as they have an extension in the carboxy-terminal part of the protein that lies on the cytoplasmic side of the membrane and may contribute to photoreceptor organization in the retina<sup>4</sup>. This extension does not affect G-protein activation, so Murakami and Kouyama<sup>2</sup> removed part of it to obtain three-dimensional crystals of the transmembrane domains. This transmembrane core was extracted from isolated photoreceptor membranes and crystallized<sup>5</sup>, the high-resolution structure being determined by X-ray crystallography. From the arrangement of the light-absorbing chromophore within the crystal packing, the authors propose a detection mechanism for linear polarized light on the principle that it is similar to the *in vivo* arrangement of the chromophore in the squid retina.

The core of squid rhodopsin consists of two cytoplasmic helices as well as seven transmembrane domains (Fig. 1). The crystal structure<sup>2</sup> shows that the arrangement of helices I to VIII is similar to that of the rhodopsins isolated from frogs and cattle, so the overall three-dimensional structure of the protein is very similar in vertebrates and invertebrates. But helices V and VI of squid rhodopsin protrude 25 Å farther into the cytoplasm from the membrane surface, with an additional helix IX contributing to the cytoplasmic part of rhodopsin. The extended helices V and VI, together with helix IX, explain the additional electron density on the cytoplasmic surface that was found in an electron-density map of two-dimensional crystals of squid



**Figure 1 | The architecture of squid rhodopsin, as determined by Murakami and Kouyama<sup>2</sup>.** The main feature of this G-protein-coupled receptor, as of all members of this superfamily of proteins, is its seven transmembrane domains, in which the polypeptide chain is organized into helices that cross the membrane. Of these, helices V and VI protrude particularly deeply into the cytoplasm of the squid photoreceptor cell, with an additional helix IX contributing to the cytoplasmic domain of the receptor. Presumably, it is these features that determine the nature of  $G_q$  binding and activation of the signalling cascade. The light-sensitive chromophore retinal (red) is covalently linked to a nitrogen atom (blue) of a lysine residue in helix VII (not labelled). A tyrosine and a glutamate residue are located close to this protonated nitrogen atom. A tryptophan residue in direct contact with retinal, as well as an internal cluster of water molecules (light blue), may be involved in transmitting the conformational change from the receptor's transmembrane core to the cytoplasmic surface.

rhodopsin<sup>4</sup>. However, the architecture of this helical dome-like structure is entirely new and could not have been predicted.

Murakami and Kouyama's structure reveals that the amino-acid residues in contact with retinal, and the orientation of retinal within the protein, are different from those in vertebrate rhodopsins. Retinal is in a less distorted configuration in squid rhodopsin. The binding pocket for it is particularly interesting, because the side chain of a glutamate residue proposed previously to provide the counterion to the positively charged nitrogen atom that provides the covalent bond to retinal is not close enough to the nitrogen atom to act in this way.

This finding has implications for the tuning of the light-absorption maximum and the mechanism of re-isomerization that returns rhodopsin to its original conformation. In this conformation, which rhodopsin assumes in the dark, the side chain of a tyrosine residue that replaces the glutamate counterion of bovine rhodopsin is close to the positively charged nitrogen on rhodopsin's lysine side chain but remains unchanged during photoactivation. This could partly explain why the covalently bound retinal is not released from the lysine side chain after photoisomerization.

As a consequence, retinal may be isomerized back to its original form within the photoreceptor to regenerate the ground state and reset the sensory system.

Several other features of the structure<sup>2</sup> merit comment. For example, an interhelical cavity containing several ordered water molecules occurs in a similar position in squid and bovine rhodopsin, as well as in vertebrate  $\beta$ -adrenergic receptors<sup>6</sup>, which also belong to the GPCR family. This water cluster may contribute to the activation mechanism by affecting the transmission of the conformational changes triggered by retinal isomerization to the G-protein-binding cytoplasmic surface<sup>7</sup>.

The arrangement of the helices, as well as the charged side chains of helices VI and IX on the cytoplasmic side, may provide a binding surface for the carboxy terminus of the  $G_q$   $\alpha$ -subunit. In addition, the extended helices V and VI indicate that  $G_q$  might bind next to these helices and not on top of the helical dome structure. This is a novel constraint for models of G-protein-receptor complexes in general, given the expected structural conservation of the activation mechanism. The unexpected architecture of this cytoplasmic surface will attract much attention. The arrangement of the helices in this domain could be used to model the receptor interaction with G proteins, as well as for designing drugs that interfere with G-protein binding.

Finally, Murakami and Kouyama's structure<sup>2</sup> will be particularly helpful for those investigating the action of human melanopsin, which — on the basis of sequence similarity — is more closely related to invertebrate than to vertebrate rhodopsin. Melanopsin is the photoreceptor that is required for setting our biological clock and for the pupillary reflex<sup>8,9</sup>, and so the structure of squid rhodopsin, in providing the best template to date for understanding this protein, is of considerable interest.

Gebhard F. X. Schertler is in the MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. e-mail: gfx@mrc-lmb.cam.ac.uk

1. Hardie, R. C. & Postma, M. in *The Senses: A Comprehensive Reference* Vol. 1 (ed. Basbaum, A. I.) 77–130 (Academic, San Diego, 2008).
2. Murakami, M. & Kouyama, T. *Nature* **453**, 363–367 (2008).
3. Hardie, R. C. & Raghupathi, P. *Nature* **413**, 186–193 (2001).
4. Davies, A. et al. *J. Mol. Biol.* **314**, 455–463 (2001).
5. Murakami, M., Kitahara, R., Gotoh, T. & Kouyama, T. *Acta Crystallogr. F* **63**, 475–479 (2007).
6. Kobilka, B. & Schertler, G. F. *Trends Pharmacol. Sci.* **29**, 79–83 (2008).
7. Li, J. et al. *J. Mol. Biol.* **343**, 1409–1438 (2004).
8. Melyan, Z., Tardif, E. E., Bellingham, J., Lucas, R. J. & Hankins, M. W. *Nature* **433**, 741–745 (2005).
9. Hankins, M. W., Peirson, S. N. & Foster, R. G. *Trends Neurosci.* **31**, 27–36 (2008).

**Competing financial interests:** declared (see online article for details).



## QUANTUM INFORMATION

# An integrated light circuit

Paul G. Kwiat

**There's a long wish list for a workable quantum computer: a viable system must be fast, compact and stable. The first integrated optical quantum logic circuits are a step in the right direction.**

Systems that use photons are one of several candidates for making a working quantum computer. One challenge (there are many) is to move beyond the current bulky experiments that involve just a few photons to the kind of stable, miniaturized circuits with very many logic gates that lie at the heart of conventional, classical computers. Writing in *Science*, Politi *et al.*<sup>1</sup> report how they have used integrated optics for the first time to demonstrate several primitive functions of quantum logic.

Quite generally, the promise of quantum computing lies in making use of the 'parallelism' and 'connectedness' of quantum-mechanical phenomena such as superposition (that objects can be in many states at once) and entanglement (that the properties of two objects can be tied to one another, despite being separated in space). These features can be used to solve information-processing tasks that would be impossible, or at least very difficult, for processors denied these quantum advantages<sup>2</sup>. Instances include problems related to factoring large numbers into their prime constituents (think  $21 = 3 \times 7$ , writ large); efficiently searching an unsorted database; and even hard problems in quantum mechanics. This last application might well end up being the most important of all, helping to form a robust framework for understanding future problems in science and technology.

How can quantum computers achieve such feats? At a basic level, every quantum processor is essentially a complex, interwoven web of nested interferometers in actual physical space (for optical approaches) or in some abstract space. An optical interferometer is made of mirrors and partially reflective, partially transmissive elements known as beam splitters. These are arranged into an optical obstacle course that an incident photon can traverse in two or more ways.

This multiplicity of paths (very) loosely corresponds to the parallel operation of a computer algorithm. By arranging elements in a certain way, one can construct logic gates for single quantum bits (qubits), preparing them in arbitrary superpositions of the logical 0 and 1. Other arrangements can entangle two qubits, so that the state of one is altered by the state of the other<sup>3</sup> (Fig. 1a).

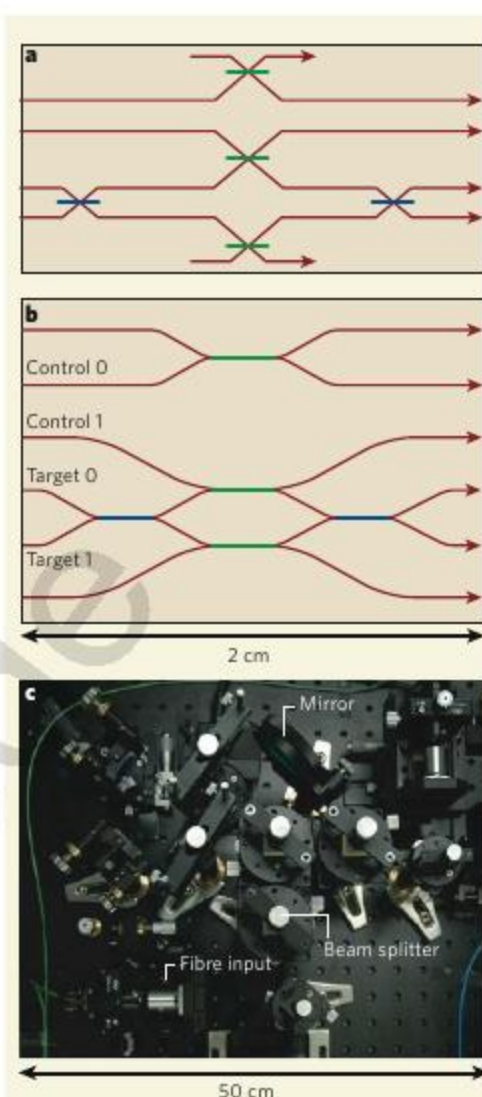
To understand where Politi and colleagues' achievement fits in, we need first to consider classical interference effects. We see these, for example, in light reflecting off a puddle with oil in it. The colours we see depend on the

direction in which we are looking, and in the details of the oil–water non-mixture (the effect depends on the varying thickness of the thin oil film lying on top of the water). This kind of swirling, shifting pattern is exactly what we don't want for a reliable computer. Experiments on quantum computing thus far have done a tolerably good job of controlling the analogous instability, and thus realizing basic quantum operations<sup>4–6</sup>.

State-of-the-art processors for optical quantum computing currently use just six photonic qubits<sup>7</sup>. But the large-scale quantum algorithms of the sort we dream of running some day (or, at least, I dream of running) will require hundreds or thousands of stable, interconnected interferometers and low-loss, high-performance components. There are ways of correcting quantum-computational errors as they arise<sup>8,9</sup>, but if they are to work properly, they require a high base-level of performance that would be nearly impossible to achieve on such a large scale. And quite apart from questions of reliability, there are also questions of speed and bulk. Just as we couldn't imagine running a modern computer application on an old-school processor that uses vacuum tubes — it would be far too bulky and slow — the complex, scaled-up quantum computer of tomorrow will require a more integrated approach<sup>10</sup>.

Enter Politi and colleagues<sup>1</sup>, with their first stab at quantum-logic operations in an (optical) integrated circuit (Fig. 1b). They succeed in replacing the bulky mirrors and protracted path lengths of traditional experiments by micrometre-scale optical waveguides — pipes for light — fabricated into a silicon wafer. Much as fibre-optic cables guide light in modern telecommunications applications, these waveguides direct the photons along their desired trajectories — rather like an irrigation system to deliver photons where and when you need them.

Such integrated optical circuits are smaller and cheaper than bulk-optic systems (or at least they will be once we have settled on an optimized design and fabrication). They are also inherently more stable; temperature variations that could alter the finely balanced path lengths in particular areas of a chip tend to cancel out. In addition, once in the waveguides, photons tend to go only where they should, so there are no optical paths to align, as there are in conventional optics. Politi *et al.* demonstrate high-quality classical and quantum interference with their system, as well as basic



**Figure 1 | Light circuits.** **a**, This circuit represents a basic two-qubit quantum logic gate. Blue beam splitters reflect 50% of the incident light, green beam splitters 33%. In this configuration, a combination of quantum and classical interference flips the target qubit states 0 and 1 if the control qubit is a 1. **b**, Politi and colleagues' first implementation<sup>1</sup> of this scheme as an integrated optical circuit uses silica-on-silicon waveguides. **c**, A bulk-optics implementation of a circuit of similar complexity involves various mirrors, beam splitters and fibre inputs; the result is considerably bigger and less stable.

one- and two-qubit logic, and entanglement with better than 92% fidelity. Their results are both a significant and a necessary step towards optical quantum computing.

Lest the reader gain an unhealthily rosy perspective, significant challenges must still be faced. There is much less opportunity to adjust these integrated circuits once they are fabricated — there is nothing to be twiddled with, unlike in conventional optical systems (Fig. 1c) — and so their design must be optimized very carefully. In addition, low-loss interconnects are a must: at present, the efficiency with which light is injected into and extracted from the circuit is 60%. That is good, but not great, and low enough to make the 99% transmission rate inside the device significantly less meaningful.

Outside the integrated circuit, other components also require work before we can achieve



large-scale quantum computing using light. For a start, we need high-quality, on-demand sources of single and entangled photons, along with a low-loss device in which to store them. Then we need photon detectors of very high efficiency (so much the better if they can actually count the number of incident photons). All of these problems are currently being attacked by various research efforts around the world. Stay tuned for more big news from tiny circuits.

Paul G. Kwiat is in the Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street,

Urbana, Illinois 61801, USA.  
e-mail: kwiat@uiuc.edu

1. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. *Science* **320**, 646–649 (2008).
2. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
3. O'Brien, J. L. *Science* **318**, 1567–1570 (2007).
4. Langford, N. K. et al. *Phys. Rev. Lett.* **95**, 210504 (2005).
5. Kiesel, N. et al. *Phys. Rev. Lett.* **95**, 210505 (2005).
6. Okamoto, R. et al. *Phys. Rev. Lett.* **95**, 210506 (2005).
7. Lu, C.-Y. et al. *Nature Phys.* **3**, 91–95 (2007).
8. Varnava, M., Browne, D. E. & Rudolph, T. *Phys. Rev. Lett.* **100**, 060502 (2008).
9. Dawson, C. M., Haselgrove, H. L. & Nielsen, M. A. *Phys. Rev. A* **73**, 052306 (2006).
10. Kielpinski, D., Monroe, C. & Wineland, D. J. *Nature* **417**, 709–711 (2002).

## TUBERCULOSIS

# Deadly combination

Stefan H. E. Kaufmann

**Many factors affect the severity of tuberculosis in infected individuals. Among these are the genetic make-up of the bacterial strain, that of the host, and the interplay between the two.**

A quick scan through this article might take five minutes of your time. During this period, 16 people will have died of tuberculosis, 80 will have fallen ill with it, and an astounding 800 will have become infected with the disease-causing pathogen *Mycobacterium tuberculosis*<sup>1</sup>. Yet, with only 1 in 10 of those infected developing tuberculosis in their lifetime, clearly most humans control the infection effectively. It is well known that resistance to tuberculosis depends on complex interactions among the host, the bacterium and the environment or culture. But the relative contributions of these factors, and the relationship between them, remain unclear. Work by Caws and colleagues<sup>2</sup>, published in *PLoS Pathogens*, is a first attempt at revealing the genetic contribution to interspecies communication between *M. tuberculosis* and its human host.

When bacteria reach human lungs — borne by tiny droplets from a cough or sneeze — they are usually engulfed and destroyed by the macrophage cells of the immune system. But *M. tuberculosis* has developed mechanisms to survive these assaults.

Detection of this pathogen by the pattern recognition receptor TLR-2 on the surface of macrophages (Fig. 1) induces a signalling cascade mediated by the TLR-2 adaptor protein TIRAP. Total activation of these primed macrophages is effected by soluble immune mediators, particularly IFN- $\gamma$ , which is secreted by white blood cells known as T cells<sup>3</sup>. Fully activated macrophages fail to eradicate *M. tuberculosis*, but restrict its growth<sup>4</sup>. Consequently, the infected individual develops latent tuberculosis, remaining healthy while harbouring dormant bacteria.

Unfortunately, the pathogens can be resus-

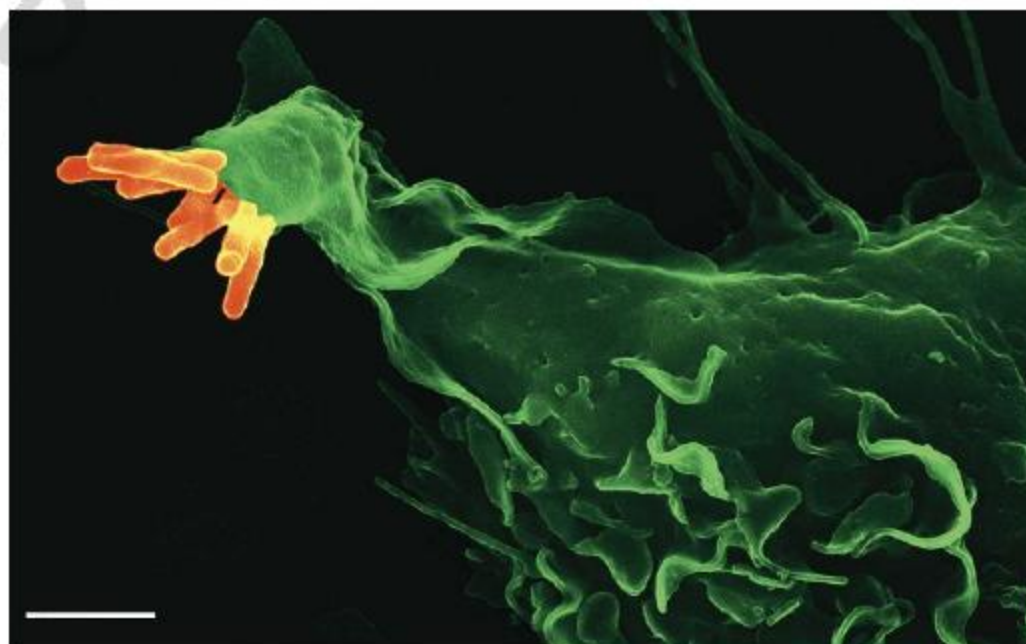
citated if the immune response diminishes, leading to pulmonary tuberculosis — the form of the disease that is confined to the lung and accounts for 80% of cases. When the bacteria reach the bronchial tree during active disease, the host becomes contagious. And once they infiltrate the bloodstream they can be dispersed to other organs. Meningeal tuberculosis, which affects the brain, is the most common (forming up to 30% of all cases) and the most hazardous form of extrapulmonary disease.

As early as the nineteenth century, it was assumed that genetic disposition contributes to the host's susceptibility to tuberculosis, but

only recently were several culprit genes identified. These genes can be divided into two groups according to whether they contribute to acquired or innate immunity. Mutations in genes involved in acquired immunity, such as the IFN- $\gamma$ -mediated signalling pathways<sup>3</sup>, are relatively rare but invariably lead to mycobacterial disease on infection — usually early in life. Genes modulating innate immunity operate more subtly, through natural genetic variations (polymorphisms), in both a synergistic and an antagonistic way. Examples of such genes include those encoding TLR-2 and TIRAP.

Polymorphisms are also rife among the different strains of *M. tuberculosis*. Different lineages of this bacterium exist that may have co-evolved in close relationship with a specific host population<sup>5</sup>. Also, separate bacterial families have developed within lineages. The Beijing family of the East Asian lineage, notorious for causing multidrug-resistant tuberculosis and spreading globally, is a noteworthy example.

Caws et al.<sup>2</sup> analysed polymorphisms both in Vietnamese adults with pulmonary or meningeal tuberculosis and in *M. tuberculosis* strains isolated from these patients. Their observations are consistent with most, although not all, previous work linking<sup>6–9</sup> polymorphisms in TLR-2/TIRAP with susceptibility to tuberculosis within the same and different ethnic groups in West Africa and Turkey. They report a close relationship between polymorphisms in the gene encoding TLR-2 and susceptibility to infection with the Beijing strain of bacterium. One particular TLR-2 polymorphism probably fails to mediate the full priming of macrophages, rendering individuals carrying it more susceptible to infection. The combination of a defective TLR-2 response and infection with the virulent Beijing strain is especially problematic, as it may increase the risk of extrapulmonary dissemination and



**Figure 1 | Host and pathogen.** On reaching the host's lungs, cells of *Mycobacterium tuberculosis* (orange) bind to the TLR-2 receptor on the surface of a macrophage (green). Caws et al.<sup>2</sup> speculate that polymorphisms both in the host genes (those encoding TLR-2 and its adaptor protein TIRAP) and in bacterial virulence factors (PGL and the DosR complex) affect the severity of tuberculosis. (Scale bar, 2  $\mu$ m.)

S. H. E. KAUFMANN & V. BRINKMANN, MPI INFECT BIOL.



large-scale quantum computing using light. For a start, we need high-quality, on-demand sources of single and entangled photons, along with a low-loss device in which to store them. Then we need photon detectors of very high efficiency (so much the better if they can actually count the number of incident photons). All of these problems are currently being attacked by various research efforts around the world. Stay tuned for more big news from tiny circuits.

Paul G. Kwiat is in the Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street,

Urbana, Illinois 61801, USA.  
e-mail: kwiat@uiuc.edu

1. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. *Science* **320**, 646–649 (2008).
2. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
3. O'Brien, J. L. *Science* **318**, 1567–1570 (2007).
4. Langford, N. K. et al. *Phys. Rev. Lett.* **95**, 210504 (2005).
5. Kiesel, N. et al. *Phys. Rev. Lett.* **95**, 210505 (2005).
6. Okamoto, R. et al. *Phys. Rev. Lett.* **95**, 210506 (2005).
7. Lu, C.-Y. et al. *Nature Phys.* **3**, 91–95 (2007).
8. Varnava, M., Browne, D. E. & Rudolph, T. *Phys. Rev. Lett.* **100**, 060502 (2008).
9. Dawson, C. M., Haselgrove, H. L. & Nielsen, M. A. *Phys. Rev. A* **73**, 052306 (2006).
10. Kielpinski, D., Monroe, C. & Wineland, D. J. *Nature* **417**, 709–711 (2002).

## TUBERCULOSIS

# Deadly combination

Stefan H. E. Kaufmann

**Many factors affect the severity of tuberculosis in infected individuals. Among these are the genetic make-up of the bacterial strain, that of the host, and the interplay between the two.**

A quick scan through this article might take five minutes of your time. During this period, 16 people will have died of tuberculosis, 80 will have fallen ill with it, and an astounding 800 will have become infected with the disease-causing pathogen *Mycobacterium tuberculosis*<sup>1</sup>. Yet, with only 1 in 10 of those infected developing tuberculosis in their lifetime, clearly most humans control the infection effectively. It is well known that resistance to tuberculosis depends on complex interactions among the host, the bacterium and the environment or culture. But the relative contributions of these factors, and the relationship between them, remain unclear. Work by Caws and colleagues<sup>2</sup>, published in *PLoS Pathogens*, is a first attempt at revealing the genetic contribution to interspecies communication between *M. tuberculosis* and its human host.

When bacteria reach human lungs — borne by tiny droplets from a cough or sneeze — they are usually engulfed and destroyed by the macrophage cells of the immune system. But *M. tuberculosis* has developed mechanisms to survive these assaults.

Detection of this pathogen by the pattern recognition receptor TLR-2 on the surface of macrophages (Fig. 1) induces a signalling cascade mediated by the TLR-2 adaptor protein TIRAP. Total activation of these primed macrophages is effected by soluble immune mediators, particularly IFN- $\gamma$ , which is secreted by white blood cells known as T cells<sup>3</sup>. Fully activated macrophages fail to eradicate *M. tuberculosis*, but restrict its growth<sup>4</sup>. Consequently, the infected individual develops latent tuberculosis, remaining healthy while harbouring dormant bacteria.

Unfortunately, the pathogens can be resus-

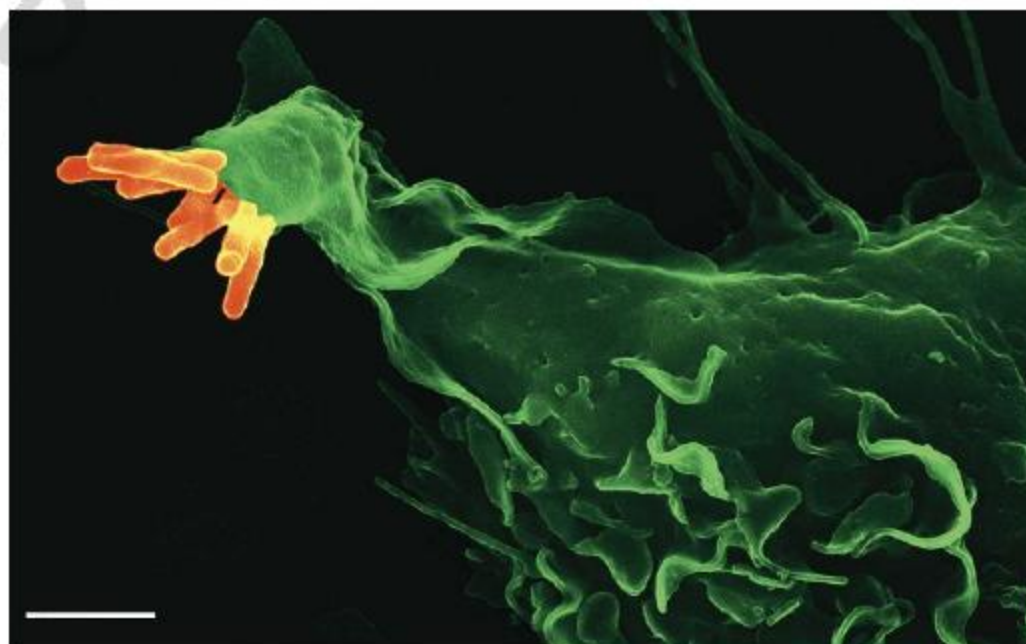
cited if the immune response diminishes, leading to pulmonary tuberculosis — the form of the disease that is confined to the lung and accounts for 80% of cases. When the bacteria reach the bronchial tree during active disease, the host becomes contagious. And once they infiltrate the bloodstream they can be dispersed to other organs. Meningeal tuberculosis, which affects the brain, is the most common (forming up to 30% of all cases) and the most hazardous form of extrapulmonary disease.

As early as the nineteenth century, it was assumed that genetic disposition contributes to the host's susceptibility to tuberculosis, but

only recently were several culprit genes identified. These genes can be divided into two groups according to whether they contribute to acquired or innate immunity. Mutations in genes involved in acquired immunity, such as the IFN- $\gamma$ -mediated signalling pathways<sup>3</sup>, are relatively rare but invariably lead to mycobacterial disease on infection — usually early in life. Genes modulating innate immunity operate more subtly, through natural genetic variations (polymorphisms), in both a synergistic and an antagonistic way. Examples of such genes include those encoding TLR-2 and TIRAP.

Polymorphisms are also rife among the different strains of *M. tuberculosis*. Different lineages of this bacterium exist that may have co-evolved in close relationship with a specific host population<sup>5</sup>. Also, separate bacterial families have developed within lineages. The Beijing family of the East Asian lineage, notorious for causing multidrug-resistant tuberculosis and spreading globally, is a noteworthy example.

Caws et al.<sup>2</sup> analysed polymorphisms both in Vietnamese adults with pulmonary or meningeal tuberculosis and in *M. tuberculosis* strains isolated from these patients. Their observations are consistent with most, although not all, previous work linking<sup>6–9</sup> polymorphisms in TLR-2/TIRAP with susceptibility to tuberculosis within the same and different ethnic groups in West Africa and Turkey. They report a close relationship between polymorphisms in the gene encoding TLR-2 and susceptibility to infection with the Beijing strain of bacterium. One particular TLR-2 polymorphism probably fails to mediate the full priming of macrophages, rendering individuals carrying it more susceptible to infection. The combination of a defective TLR-2 response and infection with the virulent Beijing strain is especially problematic, as it may increase the risk of extrapulmonary dissemination and



**Figure 1 | Host and pathogen.** On reaching the host's lungs, cells of *Mycobacterium tuberculosis* (orange) bind to the TLR-2 receptor on the surface of a macrophage (green). Caws et al.<sup>2</sup> speculate that polymorphisms both in the host genes (those encoding TLR-2 and its adaptor protein TIRAP) and in bacterial virulence factors (PGL and the DosR complex) affect the severity of tuberculosis. (Scale bar, 2  $\mu$ m.)

S. H. E. KAUFMANN & V. BRINKMANN, MPI INFECT BIOL.



severe meningeal tuberculosis. By contrast, Caws *et al.* report that the Euro-American lineage of *M. tuberculosis* mainly induces the less complicated pulmonary tuberculosis.

The Beijing strains also seem to have evolved additional virulence mechanisms to induce meningeal tuberculosis. These strains produce abundant phenolic glycolipids, which suppress innate immune responses<sup>10</sup>. Also, the activity of their DosR protein complex is increased. This complex regulates the expression of some 50 genes involved in facilitating the persistence of *M. tuberculosis* under conditions of stress<sup>11</sup> — yet another survival factor enabling Beijing strains to resist the hostile milieu of activated macrophages. Thus, infection of a person who has an impaired TLR-2/TIRAP signalling system (through polymorphisms in genes encoding these proteins) with a more virulent strain of *M. tuberculosis* such as the Beijing strains (resulting from polymorphisms that lead to better chances of bacterial survival) could be a deadly combination.

Does the improved survival of the Beijing strains mean they are more likely to lead to meningeal tuberculosis? The number of subjects Caws *et al.* examined was too small to provide any conclusive evidence, but other studies have indicated an association between members of this family of bacteria and extrapulmonary tuberculosis<sup>12</sup>. Future studies of larger subject groups should unequivocally clarify the interrelationship between *M. tuberculosis*, host genotypes and the disease characteristics.

Many years after the discovery of the first antituberculosis drug streptomycin, and despite the availability of a childhood vaccine for the disease, *M. tuberculosis* is once again emerging as a deadly pathogen across the world. The dangerous liaison between AIDS and tuberculosis continues, and incidences of multidrug-resistant, and even extensively drug-resistant, tuberculosis are on the rise. The pipeline for new tuberculosis drugs is dry, and there is no sign that a vaccine for adult tuberculosis will be available for at least a decade.

It is to be hoped, therefore, that studies such as those of Caws *et al.*<sup>2</sup> will further elucidate the multigenic interplay between humans and *M. tuberculosis*. Research will be aimed at deciphering the unknown factors underlying the transmission of *M. tuberculosis* from a patient with active tuberculosis to an uninfected individual. The development of new technologies enabling large-scale genetic screens is certainly timely. Perhaps differential gene expression will be found to contribute to both pathogen virulence and human susceptibility. Correlational studies of gene expression with analysis of functional gene products may provide definitive answers. Such studies will help to identify new drug targets and biomarkers. ■ Stefan H. E. Kaufmann is at the Max Planck Institute for Infection Biology, Department of Immunology, Charitéplatz 1, D-10117 Berlin, Germany. e-mail: kaufmann@mpiib-berlin.mpg.de

1. www.who.int/tb/publications/global\_report/en
2. Caws, M. *et al.* *PLoS Pathog.* **4**, e1000034 (2008).
3. Quintana-Murci, L., Alcais, A., Abel, L. & Casanova, J.-L. *Nature Immunol.* **8**, 1165–1171 (2007).
4. Kaufmann, S. H. E. *Nature Rev. Microbiol.* **5**, 491–504 (2007).
5. Gagneux, S. & Small, P. M. *Lancet Infect. Dis.* **7**, 328–337 (2007).
6. Khor, C. C. *et al.* *Nature Genet.* **39**, 523–528 (2007).
7. Ogus, A. C. *et al.* *Eur. Respir. J.* **23**, 219–223 (2004).
8. Hawn, T. R. *et al.* *J. Infect. Dis.* **194**, 1127–1134 (2006).
9. Nejeutsev, S. *et al.* *Nature Genet.* **40**, 261–262 (2007).
10. Reed, M. B. *et al.* *Nature* **431**, 84–87 (2004).
11. Reed, M. B. *et al.* *J. Bacteriol.* **189**, 2583–2589 (2007).
12. Kong, Y. *et al.* *J. Clin. Microbiol.* **45**, 409–414 (2007).

## CLIMATE CHANGE

# Attributing cause and effect

Francis Zwiers and Gabriele Hegerl

**The climate is changing, and so are aspects of the world's physical and biological systems. It is no easy matter to link cause and effect — the latest attack on the problem brings the power of meta-analysis to bear.**

The article by Rosenzweig and colleagues<sup>1</sup> that appears on page 353 of this issue is the first to formally link observed global changes in physical and biological systems to human-induced climate change, predominantly from increasing greenhouse gases. By surveying a huge literature, Rosenzweig *et al.* demonstrate that changes in physical and biological systems are pervasive; that these impacts lie mainly in directions consistent with warming of the climate system; and that, at least partly, they are likely to be the result of climate change caused by increasing concentrations of greenhouse gases.

The authors make the case using what is known as the 'joint attribution' approach<sup>2</sup>. They first show that the observed correspondence between impacts and warming would be very unlikely to occur if patterns of temperature change were the result of natural climate variability. They then argue that human influence has a role because observed large-scale climate change can be attributed to human influence on the climate system<sup>3</sup>.

These points emerge from Rosenzweig and colleagues' meta-analysis of the large literature on impacts, which involved synthesizing the results from studies of diverse types of system. This is probably the only available way of broadly linking impacts to climate change at global scales. Nevertheless, its very flexibility and comprehensiveness also impose limitations. For example, it would be difficult to quantify the climate-impact link with such an analysis, for two reasons.

The first is that the approach involves aggregation of results from vastly different types of system, both biological and physical (Fig. 1). The second is that only binary indicators of impacts are considered — that is, whether a given impact was consistent, or inconsistent, with warming. The link between these binary indicators and climate change is assessed by means of a technique involving a 'spatial pattern congruence' statistic, which assumes that the effects of local climate change occur locally. This measure will not fully capture

connections where biological impacts result from remote climate changes, seasonal changes or changes in temperature extremes<sup>4</sup>. So the pattern of impacts does not correlate perfectly with that of annual mean warming (see Fig. 2 of the paper<sup>1</sup> on page 355).

One of the challenges in this kind of research arises from the limitations of available data. Uncertainties result from limited and irregular sampling in space (impacts have been studied much more intensively in some regions, particularly Europe, than in others), and from the short time span of many data sets. The records mined by Rosenzweig *et al.* were primarily for the period 1970 to 2004, with the rule that at least 20 years of data were available. This is much shorter than the 50-year and sometimes 100-year data sets typically considered in studies detecting and attributing change in basic climate variables such as temperature<sup>5</sup>, surface pressure<sup>6</sup> and precipitation<sup>5,6</sup>. Such long data sets are of course few and far between, particularly for effects of climate change. But the shortness of the records used in this research<sup>1</sup> lessens the authors' ability to put changes in the context of previously observed variations, reducing the confidence in the results compared with those based on observations made over periods of a century or longer<sup>7,8</sup>.

That said, Rosenzweig and colleagues' analysis largely overcomes sampling limitations because of the sheer number of changes reported. Their synthesis is innovative and pushes detection-and-attribution research into a much broader domain than it has previously occupied. It is also a step along the road towards understanding how, and by how much, anthropogenic factors cause the observed impacts.

To estimate the size of the anthropogenic contribution, it will ultimately be necessary to undertake direct attribution of causes of change in affected systems, rather than using two-step joint attribution, in which some aspect of change in the climate system is first attributed to an external influence, and alteration in a physical or biological system is subsequently attributed to climate change. Direct attribution would



severe meningeal tuberculosis. By contrast, Caws *et al.* report that the Euro-American lineage of *M. tuberculosis* mainly induces the less complicated pulmonary tuberculosis.

The Beijing strains also seem to have evolved additional virulence mechanisms to induce meningeal tuberculosis. These strains produce abundant phenolic glycolipids, which suppress innate immune responses<sup>10</sup>. Also, the activity of their DosR protein complex is increased. This complex regulates the expression of some 50 genes involved in facilitating the persistence of *M. tuberculosis* under conditions of stress<sup>11</sup> — yet another survival factor enabling Beijing strains to resist the hostile milieu of activated macrophages. Thus, infection of a person who has an impaired TLR-2/TIRAP signalling system (through polymorphisms in genes encoding these proteins) with a more virulent strain of *M. tuberculosis* such as the Beijing strains (resulting from polymorphisms that lead to better chances of bacterial survival) could be a deadly combination.

Does the improved survival of the Beijing strains mean they are more likely to lead to meningeal tuberculosis? The number of subjects Caws *et al.* examined was too small to provide any conclusive evidence, but other studies have indicated an association between members of this family of bacteria and extrapulmonary tuberculosis<sup>12</sup>. Future studies of larger subject groups should unequivocally clarify the interrelationship between *M. tuberculosis*, host genotypes and the disease characteristics.

Many years after the discovery of the first antituberculosis drug streptomycin, and despite the availability of a childhood vaccine for the disease, *M. tuberculosis* is once again emerging as a deadly pathogen across the world. The dangerous liaison between AIDS and tuberculosis continues, and incidences of multidrug-resistant, and even extensively drug-resistant, tuberculosis are on the rise. The pipeline for new tuberculosis drugs is dry, and there is no sign that a vaccine for adult tuberculosis will be available for at least a decade.

It is to be hoped, therefore, that studies such as those of Caws *et al.*<sup>2</sup> will further elucidate the multigenic interplay between humans and *M. tuberculosis*. Research will be aimed at deciphering the unknown factors underlying the transmission of *M. tuberculosis* from a patient with active tuberculosis to an uninfected individual. The development of new technologies enabling large-scale genetic screens is certainly timely. Perhaps differential gene expression will be found to contribute to both pathogen virulence and human susceptibility. Correlational studies of gene expression with analysis of functional gene products may provide definitive answers. Such studies will help to identify new drug targets and biomarkers. ■ Stefan H. E. Kaufmann is at the Max Planck Institute for Infection Biology, Department of Immunology, Charitéplatz 1, D-10117 Berlin, Germany. e-mail: kaufmann@mpiib-berlin.mpg.de

1. www.who.int/tb/publications/global\_report/en
2. Caws, M. *et al.* *PLoS Pathog.* **4**, e1000034 (2008).
3. Quintana-Murci, L., Alcais, A., Abel, L. & Casanova, J.-L. *Nature Immunol.* **8**, 1165–1171 (2007).
4. Kaufmann, S. H. E. *Nature Rev. Microbiol.* **5**, 491–504 (2007).
5. Gagneux, S. & Small, P. M. *Lancet Infect. Dis.* **7**, 328–337 (2007).
6. Khor, C. C. *et al.* *Nature Genet.* **39**, 523–528 (2007).
7. Ogus, A. C. *et al.* *Eur. Respir. J.* **23**, 219–223 (2004).
8. Hawn, T. R. *et al.* *J. Infect. Dis.* **194**, 1127–1134 (2006).
9. Nejeutsev, S. *et al.* *Nature Genet.* **40**, 261–262 (2007).
10. Reed, M. B. *et al.* *Nature* **431**, 84–87 (2004).
11. Reed, M. B. *et al.* *J. Bacteriol.* **189**, 2583–2589 (2007).
12. Kong, Y. *et al.* *J. Clin. Microbiol.* **45**, 409–414 (2007).

## CLIMATE CHANGE

# Attributing cause and effect

Francis Zwiers and Gabriele Hegerl

**The climate is changing, and so are aspects of the world's physical and biological systems. It is no easy matter to link cause and effect — the latest attack on the problem brings the power of meta-analysis to bear.**

The article by Rosenzweig and colleagues<sup>1</sup> that appears on page 353 of this issue is the first to formally link observed global changes in physical and biological systems to human-induced climate change, predominantly from increasing greenhouse gases. By surveying a huge literature, Rosenzweig *et al.* demonstrate that changes in physical and biological systems are pervasive; that these impacts lie mainly in directions consistent with warming of the climate system; and that, at least partly, they are likely to be the result of climate change caused by increasing concentrations of greenhouse gases.

The authors make the case using what is known as the 'joint attribution' approach<sup>2</sup>. They first show that the observed correspondence between impacts and warming would be very unlikely to occur if patterns of temperature change were the result of natural climate variability. They then argue that human influence has a role because observed large-scale climate change can be attributed to human influence on the climate system<sup>3</sup>.

These points emerge from Rosenzweig and colleagues' meta-analysis of the large literature on impacts, which involved synthesizing the results from studies of diverse types of system. This is probably the only available way of broadly linking impacts to climate change at global scales. Nevertheless, its very flexibility and comprehensiveness also impose limitations. For example, it would be difficult to quantify the climate-impact link with such an analysis, for two reasons.

The first is that the approach involves aggregation of results from vastly different types of system, both biological and physical (Fig. 1). The second is that only binary indicators of impacts are considered — that is, whether a given impact was consistent, or inconsistent, with warming. The link between these binary indicators and climate change is assessed by means of a technique involving a 'spatial pattern congruence' statistic, which assumes that the effects of local climate change occur locally. This measure will not fully capture

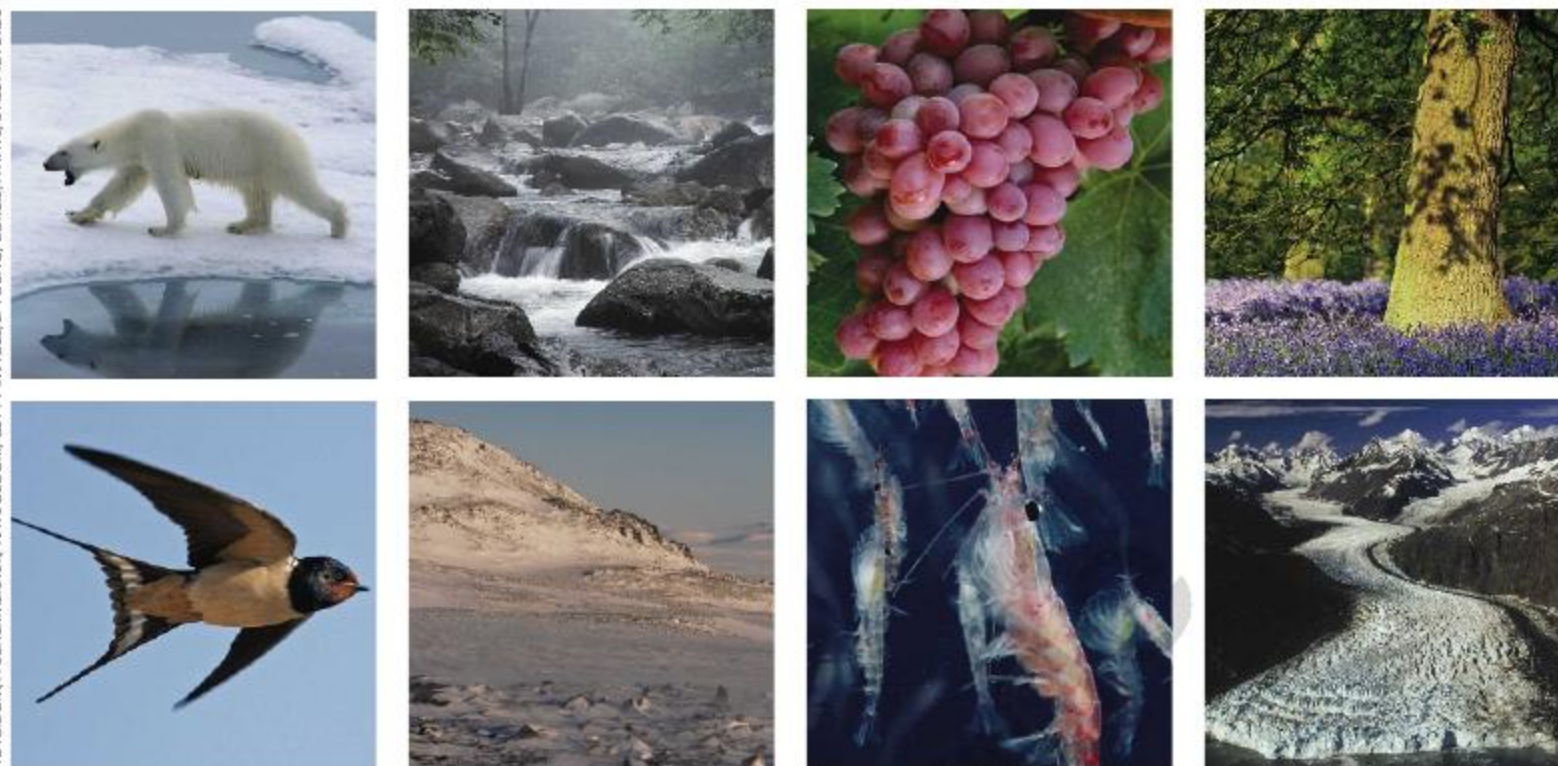
connections where biological impacts result from remote climate changes, seasonal changes or changes in temperature extremes<sup>4</sup>. So the pattern of impacts does not correlate perfectly with that of annual mean warming (see Fig. 2 of the paper<sup>1</sup> on page 355).

One of the challenges in this kind of research arises from the limitations of available data. Uncertainties result from limited and irregular sampling in space (impacts have been studied much more intensively in some regions, particularly Europe, than in others), and from the short time span of many data sets. The records mined by Rosenzweig *et al.* were primarily for the period 1970 to 2004, with the rule that at least 20 years of data were available. This is much shorter than the 50-year and sometimes 100-year data sets typically considered in studies detecting and attributing change in basic climate variables such as temperature<sup>5</sup>, surface pressure<sup>6</sup> and precipitation<sup>5,6</sup>. Such long data sets are of course few and far between, particularly for effects of climate change. But the shortness of the records used in this research<sup>1</sup> lessens the authors' ability to put changes in the context of previously observed variations, reducing the confidence in the results compared with those based on observations made over periods of a century or longer<sup>7,8</sup>.

That said, Rosenzweig and colleagues' analysis largely overcomes sampling limitations because of the sheer number of changes reported. Their synthesis is innovative and pushes detection-and-attribution research into a much broader domain than it has previously occupied. It is also a step along the road towards understanding how, and by how much, anthropogenic factors cause the observed impacts.

To estimate the size of the anthropogenic contribution, it will ultimately be necessary to undertake direct attribution of causes of change in affected systems, rather than using two-step joint attribution, in which some aspect of change in the climate system is first attributed to an external influence, and alteration in a physical or biological system is subsequently attributed to climate change. Direct attribution would





**Figure 1 | Going through the changes.** The literature sampled by Rosenzweig *et al.*<sup>1</sup> in analysing the possible impacts of climate change covered a diversity of phenomena in a diversity of biological and physical systems. Examples are perturbation of polar bear behaviour; timing of peak stream flow, of grape harvests, and of spring flowering and bird migration; and variation in the freeze-thaw pattern in tundra, in krill stocks and in glacier 'wastage'.

require an 'end-to-end'<sup>9</sup> modelling system that includes explicit representations of all of the main processes (climatic and non-climatic) that contribute to the variability of the system under study, and can simulate the response to greenhouse-gas increases as well as other factors that can cause changes in the observed impact. Such a tool can then be used to estimate how different external influences contribute to observed changes in systems relative to each other, much as models of the climate system are used to study the relative contributions of greenhouse gases, aerosols and natural climate variability to observed changes in surface air temperature and other basic climate variables<sup>3</sup>. Few such modelling systems are available, in part due to the difficulty of representing the relevant processes within climate models.

Likewise, only a few end-to-end attribution studies have been carried out<sup>1,2</sup>. They are limited to cases where the affected system and its interaction with climate are either relatively well understood<sup>10</sup> or reasonably described empirically<sup>11</sup>. We need more of them. End-to-end studies will help in interpreting the results of less direct approaches to attribution. Moreover, they can also be used to evaluate and adjust projections of future impacts based on changes that have already been observed; that will be essential in formulating strategies to adapt to the consequences of climate change, and to assess their uncertainties, much as has been done with projections of future temperature change<sup>12</sup>. The ultimate goal is to provide probabilistic projections of future effects — that is, estimates of the probability that some outcome will or will not occur — and so allow decisions

about adaptive measures to be based on a firmer footing.

Francis Zwiers is in the Climate Research Division, Environment Canada, 4905 Dufferin Street, Toronto, Ontario M3H 5T4, Canada. Gabriele Hegerl is in the School of Geosciences, University of Edinburgh, Edinburgh EH9 3JW, UK. e-mails: francis.zwiers@ec.gc.ca; gabi.hegerl@ed.ac.uk

1. Rosenzweig, C. *et al.* *Nature* **453**, 353–357 (2008).
2. Rosenzweig, C. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L. *et al.*) 79–131 (Cambridge Univ. Press, 2007).
3. Hegerl, G. C. *et al.* in *Climate Change 2007: The Physical*

*Science Basis* (eds Solomon, S. D. *et al.*) 663–745 (Cambridge Univ. Press, 2007).

4. Easterling, D. R. *et al.* *Science*, **289**, 2068–2074 (2000).
5. Zhang, X. *et al.* *Nature* **448**, 461–465 (2007).
6. Min, S.-K., Zhang, X. & Zwiers, F. *Science* **320**, 518–520 (2008).
7. Parker, D. E., Legg, T. P. & Folland, C. K. *Int. J. Clim.* **12**, 317–342 (1992).
8. Rutishauser, T., Luterbacher, J., Jeanneret, F., Pfister, C. & Wanner, H. *J. Geophys. Res.* **112**, G04016 (2007).
9. Stone, D. A. & Allen, M. R. *Climatic Change* **71**, 303–318 (2005).
10. Barnett, T. P. *et al.* *Science* **319**, 1080–1083 (2008).
11. Gillett, N. P., Weaver, A. J., Zwiers, F. W. & Flannigan, M. D. *Geophys. Res. Lett.* **31**, L18211, doi:10.1029/2004GL020876 (2004).
12. Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. D. *et al.*) 747–845 (Cambridge Univ. Press, 2007).

## SOLID-STATE PHYSICS

# Polaritronics in view

Benoît Deveaud-Plédran

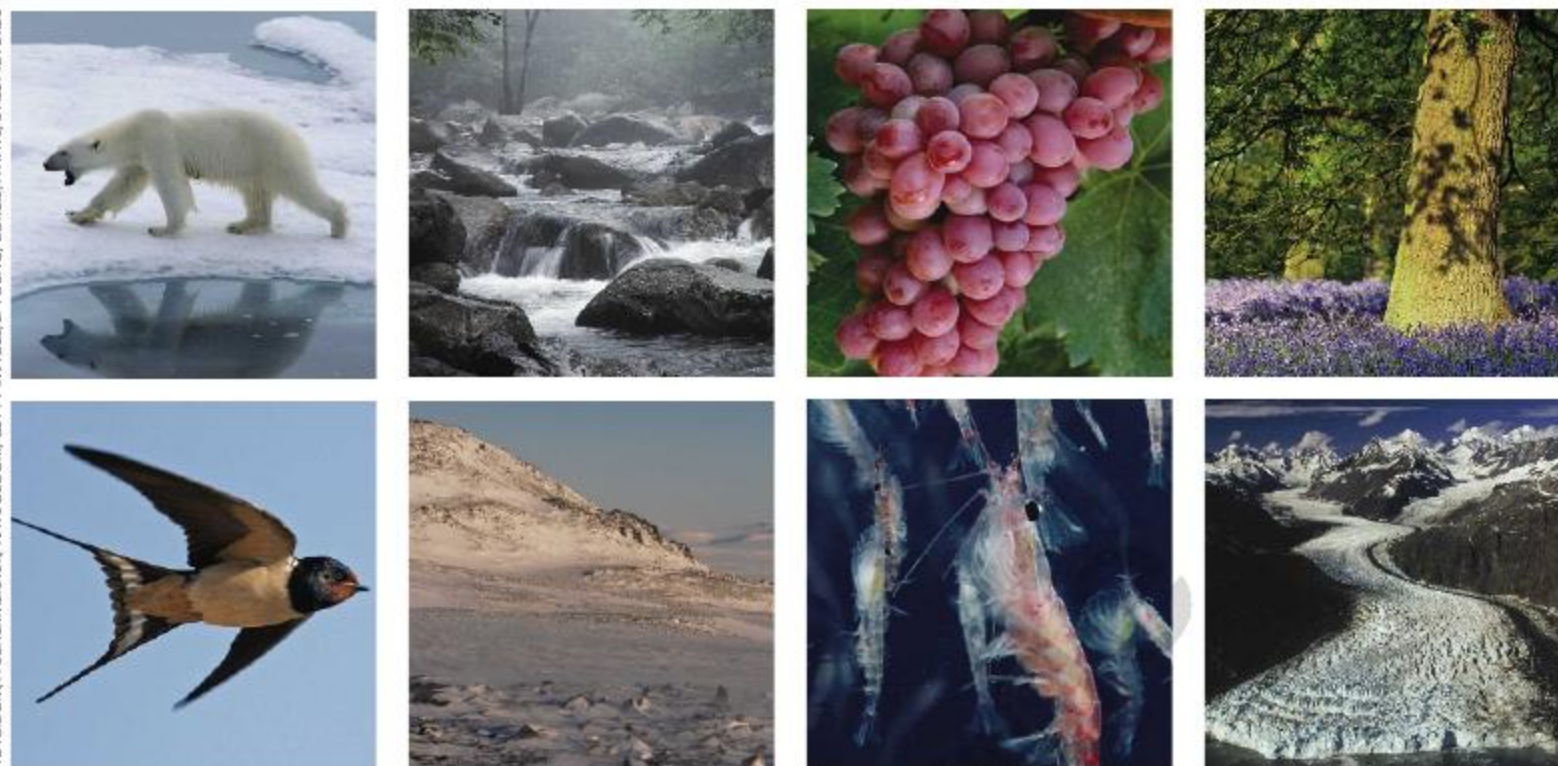
**Polaritons are an odd cross-breed of a particle, half-matter, half-light. They could offer an abundant crop of new and improved optoelectronic devices — a promise already being fulfilled.**

When I first encountered the idea of a polariton, while reading the doctoral work of John Hopfield<sup>1</sup>, its beauty stunned me. These mixed 'quasiparticles' are produced in semiconductor materials when the pairing of an electron and its phantom, a hole (this combination in itself a quasiparticle known as an exciton), couples with the photons of a light field. The result is something that is part matter, part light, and inherits qualities of both. Yet for all that they

are useful for those who wish to get to grips with the optical properties of semiconductors, I always had a feeling that polaritons were not actually 'real'. I never expected that they would make their way out of the laboratory; still less that they would ever have any practical application.

History is proving me wrong. On page 372 of this issue, Tsintzos *et al.*<sup>2</sup> detail the latest stage of the long march of polaritons to workaday





**Figure 1 | Going through the changes.** The literature sampled by Rosenzweig *et al.*<sup>1</sup> in analysing the possible impacts of climate change covered a diversity of phenomena in a diversity of biological and physical systems. Examples are perturbation of polar bear behaviour; timing of peak stream flow, of grape harvests, and of spring flowering and bird migration; and variation in the freeze-thaw pattern in tundra, in krill stocks and in glacier 'wastage'.

require an 'end-to-end'<sup>9</sup> modelling system that includes explicit representations of all of the main processes (climatic and non-climatic) that contribute to the variability of the system under study, and can simulate the response to greenhouse-gas increases as well as other factors that can cause changes in the observed impact. Such a tool can then be used to estimate how different external influences contribute to observed changes in systems relative to each other, much as models of the climate system are used to study the relative contributions of greenhouse gases, aerosols and natural climate variability to observed changes in surface air temperature and other basic climate variables<sup>3</sup>. Few such modelling systems are available, in part due to the difficulty of representing the relevant processes within climate models.

Likewise, only a few end-to-end attribution studies have been carried out<sup>1,2</sup>. They are limited to cases where the affected system and its interaction with climate are either relatively well understood<sup>10</sup> or reasonably described empirically<sup>11</sup>. We need more of them. End-to-end studies will help in interpreting the results of less direct approaches to attribution. Moreover, they can also be used to evaluate and adjust projections of future impacts based on changes that have already been observed; that will be essential in formulating strategies to adapt to the consequences of climate change, and to assess their uncertainties, much as has been done with projections of future temperature change<sup>12</sup>. The ultimate goal is to provide probabilistic projections of future effects — that is, estimates of the probability that some outcome will or will not occur — and so allow decisions

about adaptive measures to be based on a firmer footing.

Francis Zwiers is in the Climate Research Division, Environment Canada, 4905 Dufferin Street, Toronto, Ontario M3H 5T4, Canada. Gabriele Hegerl is in the School of Geosciences, University of Edinburgh, Edinburgh EH9 3JW, UK. e-mails: francis.zwiers@ec.gc.ca; gabi.hegerl@ed.ac.uk

1. Rosenzweig, C. *et al.* *Nature* **453**, 353–357 (2008).
2. Rosenzweig, C. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L. *et al.*) 79–131 (Cambridge Univ. Press, 2007).
3. Hegerl, G. C. *et al.* in *Climate Change 2007: The Physical*

*Science Basis* (eds Solomon, S. D. *et al.*) 663–745 (Cambridge Univ. Press, 2007).

4. Easterling, D. R. *et al.* *Science*, **289**, 2068–2074 (2000).
5. Zhang, X. *et al.* *Nature* **448**, 461–465 (2007).
6. Min, S.-K., Zhang, X. & Zwiers, F. *Science* **320**, 518–520 (2008).
7. Parker, D. E., Legg, T. P. & Folland, C. K. *Int. J. Clim.* **12**, 317–342 (1992).
8. Rutishauser, T., Luterbacher, J., Jeanneret, F., Pfister, C. & Wanner, H. *J. Geophys. Res.* **112**, G04016 (2007).
9. Stone, D. A. & Allen, M. R. *Climatic Change* **71**, 303–318 (2005).
10. Barnett, T. P. *et al.* *Science* **319**, 1080–1083 (2008).
11. Gillett, N. P., Weaver, A. J., Zwiers, F. W. & Flannigan, M. D. *Geophys. Res. Lett.* **31**, L18211, doi:10.1029/2004GL020876 (2004).
12. Meehl, G. A. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. D. *et al.*) 747–845 (Cambridge Univ. Press, 2007).

## SOLID-STATE PHYSICS

# Polaritronics in view

Benoît Deveaud-Plédran

**Polaritons are an odd cross-breed of a particle, half-matter, half-light. They could offer an abundant crop of new and improved optoelectronic devices — a promise already being fulfilled.**

When I first encountered the idea of a polariton, while reading the doctoral work of John Hopfield<sup>1</sup>, its beauty stunned me. These mixed 'quasiparticles' are produced in semiconductor materials when the pairing of an electron and its phantom, a hole (this combination in itself a quasiparticle known as an exciton), couples with the photons of a light field. The result is something that is part matter, part light, and inherits qualities of both. Yet for all that they

are useful for those who wish to get to grips with the optical properties of semiconductors, I always had a feeling that polaritons were not actually 'real'. I never expected that they would make their way out of the laboratory; still less that they would ever have any practical application.

History is proving me wrong. On page 372 of this issue, Tsintzos *et al.*<sup>2</sup> detail the latest stage of the long march of polaritons to workaday



respectability. They describe how they have produced a gallium arsenide diode that emits light directly from polariton states when they break up, at temperatures of up to 235 kelvin — just 60 kelvin or so below room temperature.

This breakthrough is significant, as my initial reservations about polaritons were founded largely on the fact that they can be stable only at very low temperatures and densities — under normal conditions, they tend to disintegrate, transforming themselves spontaneously into run-of-the-mill photons. In a semiconductor laser, or in a light-emitting diode (LED), temperature and density are both high. Light amplification and emission in these contexts are driven not by polariton interactions, but by the behaviour of an electron–hole plasma driven by the current flowing through the diode. It is a simple and robust mode of operation that has remained essentially unaltered in the near-50-year history of such devices, even in the newest vertical-cavity surface-emitting lasers (VCSELs). Over the years, the only factors that have changed are that the size of devices has decreased, and their efficiency has improved.

This story took a new twist in 1991, when a new type of polariton, the cavity polariton<sup>3</sup>, came to light. These quasiparticles arise when electrons and holes, in the form of excitons, are confined by the changes in the chemical composition of a substrate to a quantum energy well; simultaneously, photons are localized in the same region using two highly reflecting mirrors. In this way, stable polaritons can be produced. Stable is in this context relative; to a polariton, stability is a lifetime of more than 1 picosecond, a millionth of a millionth of a second.

More pertinently to the case in hand, stability means that cavity polaritons with zero wavevector — a quantity related to their momentum — are naturally the energetic ground state of the cavity system. This ground state will emit light precisely perpendicular to the surface of the confining sample. Subsequent studies have attested that these cavity polaritons have quite a number of useful properties. First, they have large de Broglie wavelengths of around 1 micrometre, meaning that their quantum wavefunctions are large enough to be manipulated easily, making them appealing for applications in quantum optics. Second, they might be stable at room temperature, with obvious advantages. Finally, they are good bosons, meaning that they can be parametrically amplified<sup>4</sup> (that is, split up or joined together in units of different energy, a useful technique for signal amplification), and that many of them can pile up in a given quantum state, eventually leading to the formation of the state of matter known as a Bose–Einstein condensate<sup>5</sup>.

A number of patents have been filed on the strength of these admirable qualities. These home in on possible uses for cavity polaritons in, for instance, single-photon emitters, lasers, light-emitting diodes, photodetectors and

optical switches. But at least two questions need to be addressed before such applications become reality: whether polaritons can indeed be made to operate at a sensible temperature; and whether they can be activated directly by electrical means.

Polaritons have already been confined in micrometre-sized cavities<sup>6</sup>. Tsintzos and colleagues' advance<sup>2</sup> builds on that, and fits in with a body of work<sup>7–9</sup> published in recent months. Two of these papers describe the construction of VCSEL-like cavities whose mirrors are of good enough quality to induce either electroluminescence<sup>7</sup> or the emission of laser light<sup>8</sup> in the polariton-coupling regime. Both these experiments were performed at temperatures of around 10 kelvin. A third<sup>9</sup> details the realization of a polariton LED that works at up to 100 kelvin.

The new polariton LED that operates at 235 kelvin is thus a step further towards fulfilling the promise of room-temperature, electrically driven polariton devices. The high-on simultaneous advances in the field<sup>2,6–8</sup>, on

the back of the parametric amplification<sup>4</sup> and Bose–Einstein condensation<sup>5</sup> of cavity polaritons, represent in my view the first glimpses of a grand new vista — of an expansive field that one might term 'polaritronics'. It should not be long before real, practical device applications begin to fill that panorama.

Benoît Deveaud-Plédran is in the Laboratory of Quantum Optoelectronics, Ecole Polytechnique Fédérale de Lausanne, Station 3, 1015 Lausanne, Switzerland.

e-mail: benoit.deveaud-pledrean@epfl.ch

1. Hopfield, J. J. *Phys. Rev.* **112**, 1555–1567 (1958).
2. Tsintzos, S. I., Pelekanos, N. T., Konstantinidis, G., Hatzopoulos, Z. & Savvidis, P. G. *Nature* **453**, 372–375 (2008).
3. Weisbuch, C., Nishioka, M., Ishikawa, A. & Arakawa, Y. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
4. Saba, M. *et al.* *Nature* **414**, 731–735 (2001).
5. Kasprzak, J. *et al.* *Nature* **443**, 409–414 (2006).
6. El Daif, O. *et al.* *Appl. Phys. Lett.* **88**, 061105 (2006).
7. Khalifa, A. A., Love, A. P. D., Krizhanovskii, D. N., Skolnick, M. S. & Roberts, J. S. *Appl. Phys. Lett.* **92**, 061107 (2008).
8. Bajoni, D. *et al.* *Phys. Rev. Lett.* **100**, 047401 (2008).
9. Bajoni, D. *et al.* *Phys. Rev. B* **77**, 113303 (2008).

## PLANT BIOLOGY

# In their neighbour's shadow

Jiří Friml and Michael Sauer

**They can't move away from shade, so plants resort to a molecular solution to find a place in the sun. The action they take is quite radical, and involves a reprogramming of their development.**

To survive, organisms must adapt to their ever-changing environment. Animals rely mainly on behavioural adaptive responses such as fighting or fleeing. But being stationary, plants must adjust their shape and metabolism accordingly. Plant hormones play an essential part in these adaptive responses, affecting various physiological and developmental processes. The hormone almost universally involved in plant adaptation is auxin. It exerts its effect at several developmental levels ranging from cell elongation and formation of the embryonic axis to fruit ripening<sup>1</sup>. Reporting in *Cell*, two teams — Tao *et al.*<sup>2</sup> and Stepanova *et al.*<sup>3</sup> — identify an enzyme that catalyses the first step in the biosynthesis of auxin, a step that occurs in response to changes in both ambient light and another plant hormone, ethylene.

A plant's repertoire of developmental tricks is extraordinarily broad: permanent stem-cell populations ensure growth throughout life; post-embryonic development allows new organs, such as leaves and flowers, to be generated; and differential growth enables developing plants to seek light, and roots to seek water. So even if plants have to compete — for example, for sunlight with their neighbours — they do so by modulating their own growth rather than by directly preventing that of

others. Indeed, a reduction in the quality of light causes shade-avoidance syndrome, a physiological response leading to stem elongation, fewer branches and earlier flowering.

All of these processes are mediated by plant hormones, which, like animal hormones, do not necessarily act at the location at which they are synthesized. But unlike animals, plants lack a cardiovascular system, making effective distribution of hormones problematic. Consequently, the production of plant hormones is not as localized as that of their animal counterparts, and their effect typically depends on the activation of several hormonal pathways and crosstalk between them. Individual hormonal pathways in plants have been generally well characterized at a molecular level, but research into hormone crosstalk is still in its infancy.

For auxin, spatial differences in its concentration (forming auxin gradients) are crucial for specific developmental responses<sup>1</sup>. Local manipulation of cellular auxin levels — for example, by applying auxin in droplets or by locally activating its synthesis — confirmed<sup>4,5</sup> that an increase in the level of this hormone triggers developmental programmes. So a central question in plant biology is how auxin gradients are generated. There are several answers. One is that specialized auxin-transport



respectability. They describe how they have produced a gallium arsenide diode that emits light directly from polariton states when they break up, at temperatures of up to 235 kelvin — just 60 kelvin or so below room temperature.

This breakthrough is significant, as my initial reservations about polaritons were founded largely on the fact that they can be stable only at very low temperatures and densities — under normal conditions, they tend to disintegrate, transforming themselves spontaneously into run-of-the-mill photons. In a semiconductor laser, or in a light-emitting diode (LED), temperature and density are both high. Light amplification and emission in these contexts are driven not by polariton interactions, but by the behaviour of an electron–hole plasma driven by the current flowing through the diode. It is a simple and robust mode of operation that has remained essentially unaltered in the near-50-year history of such devices, even in the newest vertical-cavity surface-emitting lasers (VCSELs). Over the years, the only factors that have changed are that the size of devices has decreased, and their efficiency has improved.

This story took a new twist in 1991, when a new type of polariton, the cavity polariton<sup>3</sup>, came to light. These quasiparticles arise when electrons and holes, in the form of excitons, are confined by the changes in the chemical composition of a substrate to a quantum energy well; simultaneously, photons are localized in the same region using two highly reflecting mirrors. In this way, stable polaritons can be produced. Stable is in this context relative; to a polariton, stability is a lifetime of more than 1 picosecond, a millionth of a millionth of a second.

More pertinently to the case in hand, stability means that cavity polaritons with zero wavevector — a quantity related to their momentum — are naturally the energetic ground state of the cavity system. This ground state will emit light precisely perpendicular to the surface of the confining sample. Subsequent studies have attested that these cavity polaritons have quite a number of useful properties. First, they have large de Broglie wavelengths of around 1 micrometre, meaning that their quantum wavefunctions are large enough to be manipulated easily, making them appealing for applications in quantum optics. Second, they might be stable at room temperature, with obvious advantages. Finally, they are good bosons, meaning that they can be parametrically amplified<sup>4</sup> (that is, split up or joined together in units of different energy, a useful technique for signal amplification), and that many of them can pile up in a given quantum state, eventually leading to the formation of the state of matter known as a Bose–Einstein condensate<sup>5</sup>.

A number of patents have been filed on the strength of these admirable qualities. These home in on possible uses for cavity polaritons in, for instance, single-photon emitters, lasers, light-emitting diodes, photodetectors and

optical switches. But at least two questions need to be addressed before such applications become reality: whether polaritons can indeed be made to operate at a sensible temperature; and whether they can be activated directly by electrical means.

Polaritons have already been confined in micrometre-sized cavities<sup>6</sup>. Tsintzos and colleagues' advance<sup>2</sup> builds on that, and fits in with a body of work<sup>7–9</sup> published in recent months. Two of these papers describe the construction of VCSEL-like cavities whose mirrors are of good enough quality to induce either electroluminescence<sup>7</sup> or the emission of laser light<sup>8</sup> in the polariton-coupling regime. Both these experiments were performed at temperatures of around 10 kelvin. A third<sup>9</sup> details the realization of a polariton LED that works at up to 100 kelvin.

The new polariton LED that operates at 235 kelvin is thus a step further towards fulfilling the promise of room-temperature, electrically driven polariton devices. The high-on simultaneous advances in the field<sup>2,6–8</sup>, on

the back of the parametric amplification<sup>4</sup> and Bose–Einstein condensation<sup>5</sup> of cavity polaritons, represent in my view the first glimpses of a grand new vista — of an expansive field that one might term 'polaritronics'. It should not be long before real, practical device applications begin to fill that panorama.

Benoît Deveaud-Plédran is in the Laboratory of Quantum Optoelectronics, Ecole Polytechnique Fédérale de Lausanne, Station 3, 1015 Lausanne, Switzerland.

e-mail: benoit.deveaud-pledrean@epfl.ch

1. Hopfield, J. J. *Phys. Rev.* **112**, 1555–1567 (1958).
2. Tsintzos, S. I., Pelekanos, N. T., Konstantinidis, G., Hatzopoulos, Z. & Savvidis, P. G. *Nature* **453**, 372–375 (2008).
3. Weisbuch, C., Nishioka, M., Ishikawa, A. & Arakawa, Y. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
4. Saba, M. *et al.* *Nature* **414**, 731–735 (2001).
5. Kasprzak, J. *et al.* *Nature* **443**, 409–414 (2006).
6. El Daif, O. *et al.* *Appl. Phys. Lett.* **88**, 061105 (2006).
7. Khalifa, A. A., Love, A. P. D., Krizhanovskii, D. N., Skolnick, M. S. & Roberts, J. S. *Appl. Phys. Lett.* **92**, 061107 (2008).
8. Bajoni, D. *et al.* *Phys. Rev. Lett.* **100**, 047401 (2008).
9. Bajoni, D. *et al.* *Phys. Rev. B* **77**, 113303 (2008).

## PLANT BIOLOGY

# In their neighbour's shadow

Jiří Friml and Michael Sauer

**They can't move away from shade, so plants resort to a molecular solution to find a place in the sun. The action they take is quite radical, and involves a reprogramming of their development.**

To survive, organisms must adapt to their ever-changing environment. Animals rely mainly on behavioural adaptive responses such as fighting or fleeing. But being stationary, plants must adjust their shape and metabolism accordingly. Plant hormones play an essential part in these adaptive responses, affecting various physiological and developmental processes. The hormone almost universally involved in plant adaptation is auxin. It exerts its effect at several developmental levels ranging from cell elongation and formation of the embryonic axis to fruit ripening<sup>1</sup>. Reporting in *Cell*, two teams — Tao *et al.*<sup>2</sup> and Stepanova *et al.*<sup>3</sup> — identify an enzyme that catalyses the first step in the biosynthesis of auxin, a step that occurs in response to changes in both ambient light and another plant hormone, ethylene.

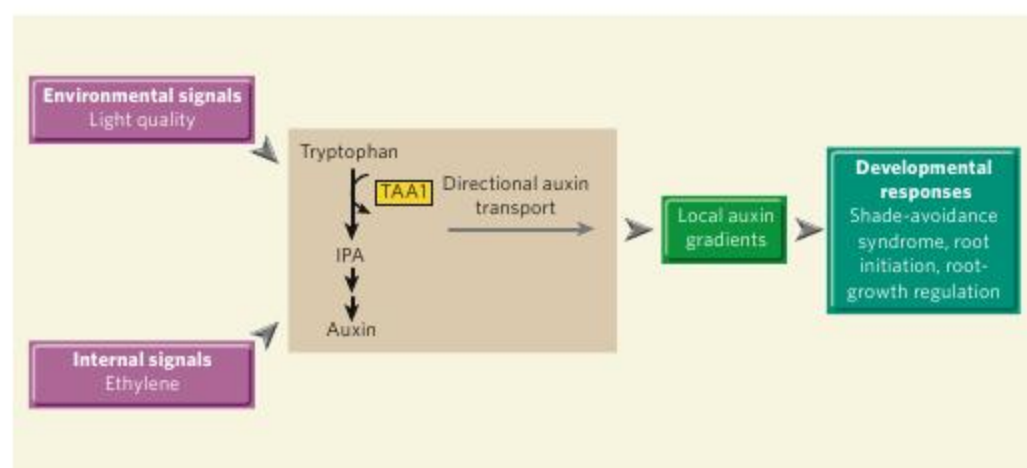
A plant's repertoire of developmental tricks is extraordinarily broad: permanent stem-cell populations ensure growth throughout life; post-embryonic development allows new organs, such as leaves and flowers, to be generated; and differential growth enables developing plants to seek light, and roots to seek water. So even if plants have to compete — for example, for sunlight with their neighbours — they do so by modulating their own growth rather than by directly preventing that of

others. Indeed, a reduction in the quality of light causes shade-avoidance syndrome, a physiological response leading to stem elongation, fewer branches and earlier flowering.

All of these processes are mediated by plant hormones, which, like animal hormones, do not necessarily act at the location at which they are synthesized. But unlike animals, plants lack a cardiovascular system, making effective distribution of hormones problematic. Consequently, the production of plant hormones is not as localized as that of their animal counterparts, and their effect typically depends on the activation of several hormonal pathways and crosstalk between them. Individual hormonal pathways in plants have been generally well characterized at a molecular level, but research into hormone crosstalk is still in its infancy.

For auxin, spatial differences in its concentration (forming auxin gradients) are crucial for specific developmental responses<sup>1</sup>. Local manipulation of cellular auxin levels — for example, by applying auxin in droplets or by locally activating its synthesis — confirmed<sup>4,5</sup> that an increase in the level of this hormone triggers developmental programmes. So a central question in plant biology is how auxin gradients are generated. There are several answers. One is that specialized auxin-transport





**Figure 1 | Plant plasticity.** Two studies<sup>2,3</sup> show that environmental cues (light quality) and internal signals (mediated by the hormone ethylene) can both modulate the auxin synthetic pathway that starts with the conversion of tryptophan to indole-3-pyruvic acid (IPA) in a reaction catalysed by the enzyme TAA1. TAA1-dependent local synthesis of auxin, together with the resultant directional auxin transport, leads to auxin gradients that mediate specific adaptive developmental responses in the plant.

proteins mediate directional transport of the hormone. Asymmetric subcellular localization of these transporters ensures directional auxin movement between cells and can generate uneven auxin distribution<sup>6</sup>. Spatially restricted peaks in auxin levels can also be generated by local auxin biosynthesis<sup>7</sup>, although the biochemical pathways involved in this process are complex and remain poorly understood. Tao *et al.*<sup>2</sup> and Stepanova *et al.*<sup>3</sup> now identify an enzyme that is essential for local auxin biosynthesis and that markedly affects plant development.

Tao and colleagues performed a genetic screen to identify mutations in the plant *Arabidopsis thaliana* that compromise shade-avoidance responses. They detected the *TAA1* gene, which when mutated prevents plant elongation under shady conditions. This mutant also has lower auxin levels and, in contrast to its normal counterpart, does not produce more auxin when placed in shade. Supplying external auxin restores the shade-avoidance response in *taa1* mutants.

The authors' elaborate biochemical and structural analyses of TAA1 convincingly show that this enzyme catalyses conversion of the amino acid tryptophan to an auxin precursor, indole-3-pyruvic acid (Fig. 1). They also report that the expression of TAA1 is localized, being predominantly confined to the leaf margins, and that it increases under shady conditions. This finding is in line with the observed defects in shade-avoidance syndrome seen in *taa1* mutants.

In addition to confirming the role of TAA1 in auxin biosynthesis, Stepanova and colleagues<sup>3</sup> investigated the developmental relevance of this enzyme. They find two related genes that partially compensate for the loss of TAA1 function. But if all three genes are inactive, the consequences are disastrous: triple mutants are severely defective even during embryonic development, failing to generate a root. Only in certain combinations of the three mutations do plants survive, but

at all developmental stages they still show strong, auxin-deficiency-related characteristics that are similar to the effect of mutations in *YUCCA* genes. (*YUCCA* genes encode enzymes that are also implicated in local auxin biosynthesis<sup>7</sup>.)

In a slight twist to the story, Stepanova *et al.* identify TAA1 in a screen for tissue-specific modulators of response to another essential plant hormone, ethylene. The *taa1* mutants were insensitive to ethylene, but external auxin application restored their sensitivity, confirming earlier speculation<sup>8–10</sup> about crosstalk between auxin and ethylene.

Taken together, these two papers<sup>2,3</sup> unambiguously identify a central component of auxin biosynthesis and demonstrate its importance for specific developmental responses, as well as for hormone crosstalk. From a biochemical perspective, it will be interesting to see where, in relation to *YUCCA*, TAA1 resides in the auxin biosynthetic pathway, as the partially overlapping characteristics of these proteins' mutants indicate a common branch. Understanding the spatial and temporal regulation of TAA1, and expression of its related genes under different environmental conditions, is also of great interest for the exciting prospect of unravelling the mechanisms of signalling crosstalk in plants.

Jiří Friml is in the Department of Plant Systems Biology, VIB, and the Department of Molecular Genetics, Ghent University, 9052 Ghent, Belgium. Michael Sauer is at the Centro Nacional de Biotecnología, Genética Molecular de Plantas, 28049 Madrid, Spain.

1. Tanaka, H. *et al.* *Cell Mol. Life Sci.* **63**, 2738–2754 (2006).
2. Tao, Y. *et al.* *Cell* **133**, 164–176 (2008).
3. Stepanova, A. N. *et al.* *Cell* **133**, 177–191 (2008).
4. Reinhardt, D. *et al.* *Nature* **426**, 255–260 (2003).
5. Dubrowski, J. *et al.* *Proc. Natl Acad. Sci. USA* (in the press).
6. Vieten, A. *et al.* *Trends Plant Sci.* **12**, 160–168 (2007).
7. Cheng, Y. *et al.* *Genes Dev.* **20**, 1790–1799 (2006).
8. Swarup, R. *et al.* *Plant Cell* **19**, 2186–2196 (2007).
9. Stepanova, A. *et al.* *Plant Cell* **19**, 2169–2185 (2007).
10. Růžička, K. *et al.* *Plant Cell* **19**, 2197–2212 (2007).



## 50 YEARS AGO

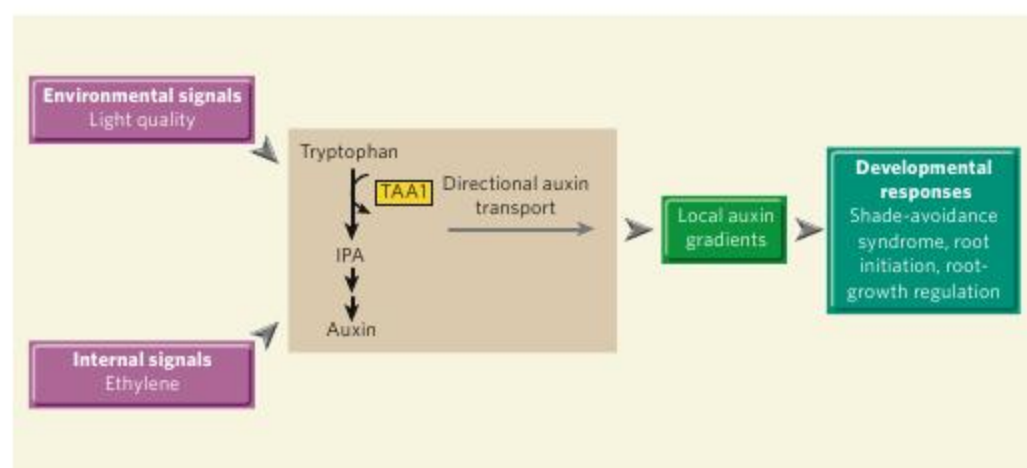
Niels Henrik Abel: Mathematician Extraordinary — ... [Abel] was a very nineteenth-century genius. The son of a drunken Norwegian clergyman and his profligate wife, he faced poverty as a student; his subsequent rise to fame was hampered by the indifference, and sometimes the misunderstanding, of the established great; his hope of financial reward was frustrated by official vulnerability to pressure groups; and he died of consumption at the age of twenty-six ... The young genius of to-day will find difficulty not in getting his work published, but in getting it noticed amid the flood from his ordinary sound contemporaries. If he dies at twenty-six, it will not be of consumption, but in a mountaineering accident; and however real his genius, so dying, he will surely not leave so deep or so lasting a mark on so wide a sector of his subject as did Abel. From *Nature* 17 May 1958.

## 100 YEARS AGO

"The Daylight Saving Bill" ... proposes that early on the morning of each of the first four Sundays in April all the public clocks shall be set forward twenty minutes and be set back twenty minutes on each of the four Sunday mornings in September ... But Sir David Gill shows, in a letter in Tuesday's *Times*, that even to change the origin of time once for all requires careful preparation, and that to make changes in the manner proposed by the Bill must lead to confusion ... He points out that if, for example, the Bank of England could be persuaded to open business at 9 a.m. instead of 10 a.m. from April 1 to the end of September, no doubt all other banks and offices would follow suit, and if employers of labour would open their works an hour earlier in the spring and summer months the objects of the Bill would be in great part gained without difficulty and confusion. From *Nature* 14 May 1908.

50 & 100 YEARS AGO





**Figure 1 | Plant plasticity.** Two studies<sup>2,3</sup> show that environmental cues (light quality) and internal signals (mediated by the hormone ethylene) can both modulate the auxin synthetic pathway that starts with the conversion of tryptophan to indole-3-pyruvic acid (IPA) in a reaction catalysed by the enzyme TAA1. TAA1-dependent local synthesis of auxin, together with the resultant directional auxin transport, leads to auxin gradients that mediate specific adaptive developmental responses in the plant.

proteins mediate directional transport of the hormone. Asymmetric subcellular localization of these transporters ensures directional auxin movement between cells and can generate uneven auxin distribution<sup>6</sup>. Spatially restricted peaks in auxin levels can also be generated by local auxin biosynthesis<sup>7</sup>, although the biochemical pathways involved in this process are complex and remain poorly understood. Tao *et al.*<sup>2</sup> and Stepanova *et al.*<sup>3</sup> now identify an enzyme that is essential for local auxin biosynthesis and that markedly affects plant development.

Tao and colleagues performed a genetic screen to identify mutations in the plant *Arabidopsis thaliana* that compromise shade-avoidance responses. They detected the *TAA1* gene, which when mutated prevents plant elongation under shady conditions. This mutant also has lower auxin levels and, in contrast to its normal counterpart, does not produce more auxin when placed in shade. Supplying external auxin restores the shade-avoidance response in *taa1* mutants.

The authors' elaborate biochemical and structural analyses of TAA1 convincingly show that this enzyme catalyses conversion of the amino acid tryptophan to an auxin precursor, indole-3-pyruvic acid (Fig. 1). They also report that the expression of TAA1 is localized, being predominantly confined to the leaf margins, and that it increases under shady conditions. This finding is in line with the observed defects in shade-avoidance syndrome seen in *taa1* mutants.

In addition to confirming the role of TAA1 in auxin biosynthesis, Stepanova and colleagues<sup>3</sup> investigated the developmental relevance of this enzyme. They find two related genes that partially compensate for the loss of TAA1 function. But if all three genes are inactive, the consequences are disastrous: triple mutants are severely defective even during embryonic development, failing to generate a root. Only in certain combinations of the three mutations do plants survive, but

at all developmental stages they still show strong, auxin-deficiency-related characteristics that are similar to the effect of mutations in *YUCCA* genes. (*YUCCA* genes encode enzymes that are also implicated in local auxin biosynthesis<sup>7</sup>.)

In a slight twist to the story, Stepanova *et al.* identify TAA1 in a screen for tissue-specific modulators of response to another essential plant hormone, ethylene. The *taa1* mutants were insensitive to ethylene, but external auxin application restored their sensitivity, confirming earlier speculation<sup>8–10</sup> about crosstalk between auxin and ethylene.

Taken together, these two papers<sup>2,3</sup> unambiguously identify a central component of auxin biosynthesis and demonstrate its importance for specific developmental responses, as well as for hormone crosstalk. From a biochemical perspective, it will be interesting to see where, in relation to *YUCCA*, TAA1 resides in the auxin biosynthetic pathway, as the partially overlapping characteristics of these proteins' mutants indicate a common branch. Understanding the spatial and temporal regulation of TAA1, and expression of its related genes under different environmental conditions, is also of great interest for the exciting prospect of unravelling the mechanisms of signalling crosstalk in plants.

Jiří Friml is in the Department of Plant Systems Biology, VIB, and the Department of Molecular Genetics, Ghent University, 9052 Ghent, Belgium. Michael Sauer is at the Centro Nacional de Biotecnología, Genética Molecular de Plantas, 28049 Madrid, Spain.

1. Tanaka, H. *et al.* *Cell Mol. Life Sci.* **63**, 2738–2754 (2006).
2. Tao, Y. *et al.* *Cell* **133**, 164–176 (2008).
3. Stepanova, A. N. *et al.* *Cell* **133**, 177–191 (2008).
4. Reinhardt, D. *et al.* *Nature* **426**, 255–260 (2003).
5. Dubrowski, J. *et al.* *Proc. Natl Acad. Sci. USA* (in the press).
6. Vieten, A. *et al.* *Trends Plant Sci.* **12**, 160–168 (2007).
7. Cheng, Y. *et al.* *Genes Dev.* **20**, 1790–1799 (2006).
8. Swarup, R. *et al.* *Plant Cell* **19**, 2186–2196 (2007).
9. Stepanova, A. *et al.* *Plant Cell* **19**, 2169–2185 (2007).
10. Růžička, K. *et al.* *Plant Cell* **19**, 2197–2212 (2007).



## 50 YEARS AGO

Niels Henrik Abel: Mathematician Extraordinary — ... [Abel] was a very nineteenth-century genius. The son of a drunken Norwegian clergyman and his profligate wife, he faced poverty as a student; his subsequent rise to fame was hampered by the indifference, and sometimes the misunderstanding, of the established great; his hope of financial reward was frustrated by official vulnerability to pressure groups; and he died of consumption at the age of twenty-six ... The young genius of to-day will find difficulty not in getting his work published, but in getting it noticed amid the flood from his ordinary sound contemporaries. If he dies at twenty-six, it will not be of consumption, but in a mountaineering accident; and however real his genius, so dying, he will surely not leave so deep or so lasting a mark on so wide a sector of his subject as did Abel. From *Nature* 17 May 1958.

## 100 YEARS AGO

"The Daylight Saving Bill" ... proposes that early on the morning of each of the first four Sundays in April all the public clocks shall be set forward twenty minutes and be set back twenty minutes on each of the four Sunday mornings in September ... But Sir David Gill shows, in a letter in Tuesday's *Times*, that even to change the origin of time once for all requires careful preparation, and that to make changes in the manner proposed by the Bill must lead to confusion ... He points out that if, for example, the Bank of England could be persuaded to open business at 9 a.m. instead of 10 a.m. from April 1 to the end of September, no doubt all other banks and offices would follow suit, and if employers of labour would open their works an hour earlier in the spring and summer months the objects of the Bill would be in great part gained without difficulty and confusion. From *Nature* 14 May 1908.

50 & 100 YEARS AGO



## OBITUARY

## Edward N. Lorenz (1917–2008)

Meteorologist and father of chaos theory.

Edward Norton Lorenz, whose pioneering studies of atmospheric dynamics led to his accidental discovery of chaos theory, died of cancer at his home in Cambridge, Massachusetts, on 16 April. A modest, unassuming and kind man, his personal qualities and intellectual insights had been a constant feature in the field of meteorology for more than 60 years; he co-authored his last paper just weeks before his death.

Born on 23 May 1917 in West Hartford, Connecticut, Lorenz took bachelor's and master's degrees in mathematics at Dartmouth College, New Hampshire, and Harvard University, respectively. Service as a weather forecaster for the US Army Air Corps during the Second World War led him into meteorology, and he received a doctorate in the subject at the Massachusetts Institute of Technology (MIT) in 1948. He remained in MIT's Department of Meteorology for the rest of his academic career, becoming emeritus professor there in 1987.

Lorenz made crucial contributions to atmospheric science, many of which are still routinely taught to students and widely used in weather forecasting. Perhaps foremost among these is his formulation in the mid-1950s of the concept of 'available potential energy', which he used to explain how potential energy and kinetic energy are interchanged in the atmosphere. His application of these ideas culminated in his influential book of 1967, *The Nature and Theory of the General Circulation of the Atmosphere*. He was also instrumental in the development of numerical techniques for weather prediction. One example — again, still widely used — is his scheme for the numerical treatment of changes in atmospheric variables with height, now known as the Lorenz vertical grid.

But the work for which Lorenz is undoubtedly most widely known is a now-classic paper published in the *Journal of Atmospheric Science* in 1963. Entitled 'Deterministic nonperiodic flow', it presented surprising results from a simplified computational model that simulated thermal convection in a fluid layer heated from below and cooled from the top. The calculated flow of the fluid was extremely irregular, with almost random qualities. But more importantly, it exhibited extremely sensitive dependence on initial conditions: two fluid states that were at first just slightly different diverged from each other exponentially, with their differences doubling repeatedly



at a consistent rate. When Lorenz plotted variables representing temperature and flow against one another, the system eventually adopted trajectories that traced out something akin to a pair of butterfly wings — a pattern since called the Lorenz attractor. He further observed that the system trajectory moved from one wing of the butterfly to another in a seemingly erratic manner.

In his book *The Essence of Chaos*, Lorenz recounts how he came to discover the extreme sensitivity of his model to small changes. Wishing to repeat his simulation, he restarted it with numbers that had been printed out for the start conditions, and left it to go down the hall to fetch a cup of coffee. On his return, he found that the result was nothing like the previous one. He soon identified the reason: the numbers from the print-out were rounded off. In the course of a coffee break, that small error had propagated with exponential speed to change the result completely.

This discovery was epoch-making for two reasons. The first lay in Lorenz's integration of analytical methods with computational simulations, with which he — albeit with a pre-1960 computer that was bulkier, noisier and vastly slower than the PCs of today — set an early precedent for a mode of research that has since become a norm. But much more profound ramifications stemmed from Lorenz's realization of just how general the types of motion he had uncovered were in

nonlinear systems such as the atmosphere. Most immediately for Lorenz's field, this meant that long-term weather predictions were impossible, because the atmosphere's initial state can never be specified precisely enough. That was a situation that increased computing power could not change.

Lorenz perfectly encapsulated this unknowability in the title of a talk that he gave to the American Association for the Advancement of Science in 1972. The question it asks has since lodged itself in the public's consciousness: "Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?" But the influence of chaos theory extends far beyond meteorology, and much deeper: it challenges the entire deterministic world view, as was confidently expressed, for instance, by the mathematician and philosopher Pierre-Simon Laplace, who stated at the beginning of the nineteenth century that the entire future could be determined by constructing and solving the equations governing all components of the Universe.

Although the existence of chaos had been recognized before Lorenz — notably in the 1890s by Henri Poincaré, in his study of the motions of three or more gravitating celestial bodies — it was Lorenz's meteorological demonstration and analysis that established the universal applicability of the concept, and earned him the title 'the father of chaos'. But it took a decade for chaos theory to percolate through to the general scientific community. When it finally did, it launched a revolution, rapidly extending its sway into many fields of physics, chemistry, biology and engineering — and, in doing so, becoming part of the popular lexicon.

Lorenz received many honours and prizes in recognition of his work, among them the Crafoord Prize — established by the Royal Swedish Academy of Sciences to recognize work in fields not covered by the Nobel prizes — in 1983, and the Kyoto Prize in 1991. The citation for that prize lauded "his boldest scientific achievement in discovering 'deterministic chaos', a principle that has profoundly influenced a wide range of basic sciences and brought about one of the most dramatic changes in mankind's view of nature since Sir Isaac Newton".

#### Edward Ott

Edward Ott is in the Departments of Physics, and of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742, USA.  
e-mail: edott@umd.edu

MIT



**Cover illustration**

A colony of human embryonic stem cells (light blue) growing on fibroblasts (dark blue). (Courtesy of A. Michalska and A. Trounson, Monash University (MISCL))

**Editor, Nature**

Philip Campbell

**Insights Publisher**

Steven Inchcoombe

**Insights Editor**

Ritu Dhand

**Production Editor**

Davina Dudley-Moore

**Senior Art Editor**

Martin Harrison

**Art Editor**

Nik Spencer

**Sponsorship**

Amélie Pequignot

**Production**

Jocelyn Hilton

**Marketing**

Katy Dunningham  
Elena Woodstock

**Editorial Assistant**

Alison McGill

# REGENERATIVE MEDICINE

**L**ife is regenerative, by definition. But by and large, humans lack the regenerative capacity of creatures such as newts and hydra. Although some of our cells have the innate ability to replenish themselves — and, by doing so, to repair ageing and injured tissues and organs — most of the body's cells form the specialized cell type they are destined for and then go into lock down.

Having said that, humans do have organs and tissues, such as liver and skin, that regenerate well. Unfortunately, the insults of injury, disease and age wreak havoc on those that don't. This explains why diseases of the heart, an organ famously recalcitrant to regeneration, are killers. And even organs that can regenerate eventually succumb to the ravages of ageing.

The field of reprogramming began with John Gurdon's seminal work on the reprogramming of frog cells by cloning. His experiments showed that somatic cells that normally cannot regenerate (the majority of cells) can be stimulated to do so in certain circumstances. By bathing the nuclei of somatic cells in protein factors obtained from eggs, or even by inserting a few genes, the cells take on the quality of embryonic cells — the most regenerative cells of all.

Human musings on regeneration are ancient, as illustrated by the tale of the luckless Greek titan Prometheus. By day, an eagle torments him by tearing out his liver, only for the organ to regenerate overnight ready to be torn out again the next day. Now, scientists are discovering our bodies' innate stem cells and how to create new sources of such cells in a Petri dish. This knowledge is transforming biology.

The articles in this Insight explore the promises and challenges of the next era of regenerative medicine — and how to use the information gained from the study of model organisms and cell culture to eventually heal ourselves. For an additional perspective, see *Nature Reports Stem Cells* ([www.nature.com/stemcells/index.html](http://www.nature.com/stemcells/index.html)) for a series of Q&As with the Insight authors.

Natalie DeWitt, Senior Editor

## INTRODUCTION

### 302 Regenerative medicine and human models of human disease

K. R. Chien

## REVIEWS

### 306 Intrinsic and extrinsic control of haematopoietic stem-cell self-renewal

L. I. Zon

### 314 Wound repair and regeneration

G. C. Gurtner, S. Werner, Y. Barrandon & M. T. Longaker

### 322 Stem-cell-based therapy and lessons from the heart

R. Passier, L. W. van Laake & C. L. Mummery

### 330 Tolerance strategies for stem-cell-based therapies

A. P. Chidgey, D. Layton, A. Trounson & R. L. Boyd

### 338 A chemical approach to stem-cell biology and regenerative medicine

Y. Xu, Y. Shi & S. Ding

### 345 Imaging stem-cell-driven regeneration in mammals

T. Schroeder

nature  
insight



# Regenerative medicine and human models of human disease

Kenneth R. Chien<sup>1,2</sup>

**Recent advances in stem-cell technology are now allowing the mechanisms of human disease to be studied in human cells. A new era for regenerative medicine is arising from such disease models, extending beyond early cell-based therapies and towards evaluating genetic variation in humans and identifying the molecular pathways that lead to disease, as well as targets for therapy.**

*Doctors are men who prescribe medicines of which they know little, to cure diseases of which they know less, in human beings of whom they know nothing*

Voltaire

Medicine has come a long way since the Age of Enlightenment. Targeted therapies and genetically engineered mouse models of human diseases are in common use, and the genome sequences of many species are known. But, despite this progress, there are still innumerable unmet clinical needs, particularly for treating degenerative diseases. For advanced forms of such diseases, it is important not only to slow the progression of disease but also to reverse its course, a key precept of regenerative medicine. But these processes are difficult without understanding the underlying disease mechanisms. Clues about the molecular pathways that lead to cardiovascular disease, for example, are rapidly being gathered from genome-wide association studies (which look for genetic variants associated with particular diseases), but for most diseases, a clear understanding of how such pathways affect disease pathogenesis is lacking. Thus, knowledge of the mechanistic 'logic' of many degenerative diseases in humans and how best to disrupt the progression of these diseases, let alone reverse them, remains at a relatively primitive stage. So far, heart transplantation is one of only a handful of successful examples of regenerative medicine, and (although several treatments are in clinical trials; Table 1) it is the only successful example of regenerative cardiovascular medicine (Box 1).

Clearly, a better understanding of disease mechanisms is required for regenerative medicine to succeed. But most of what is known about the molecular pathways that lead to human diseases has come from studying animal models, particularly genetically engineered mouse models, which do not necessarily mimic human physiology or precisely recapitulate human disease. The cardiovascular systems of mice and humans, for example, diverge with respect to serum lipid profiles, the way that abnormalities in heart rhythm develop and the phenotype that results from certain single-gene disorders. Studies using human cells, however, have been hampered by the difficulty of isolating the cell types of interest from intact human tissue.

Now, researchers have the ability to generate and expand various types of differentiated cell from both human embryonic stem (ES) cells<sup>1</sup> and human induced pluripotent stem (iPS) cells<sup>2,3</sup> (the latter of which are created by eliciting the expression of a few specific genes in adult cells, commonly skin fibroblasts, and thereby inducing them to become ES-cell-like). These technological advances provide the opportunity to set up human models of human disease, which should markedly improve the understanding of human disease mechanisms. By isolating these cells with the potential to become multiple cell types (known as multipotent cells or pluripotent cells) from individuals with particular diseases, it will be possible to generate human cardiovascular cells and tissue and therefore to study the molecular pathways that mediate disease directly in 'normal' and 'diseased' cells. It will also be possible to integrate findings from these models with higher-throughput genetic and chemical screening technologies and to obtain rare populations of human cells in larger quantities than is possible at present. In short, progress in stem-cell technology is beginning to offer the unprecedented possibility of using humans as model organisms. Moreover, as the molecular pathways that direct the differentiation of ES-cell-derived cell types are identified, this knowledge will undoubtedly enable improvements to be made to cell-based therapy for a host of degenerative diseases.

The articles in this Insight explore recent scientific advances that are likely to have a major impact on regenerative medicine. And in this introduction to the Insight, I discuss some of the challenges to constructing human disease models based on human ES cells or iPS cells, as well as some of the benefits of the cell-based therapy that these models might eventually allow, using cardiovascular disease as an example.

## The challenges

A central challenge to the development of human stem-cell-based models of disease lies in the need to isolate and expand rare cell populations

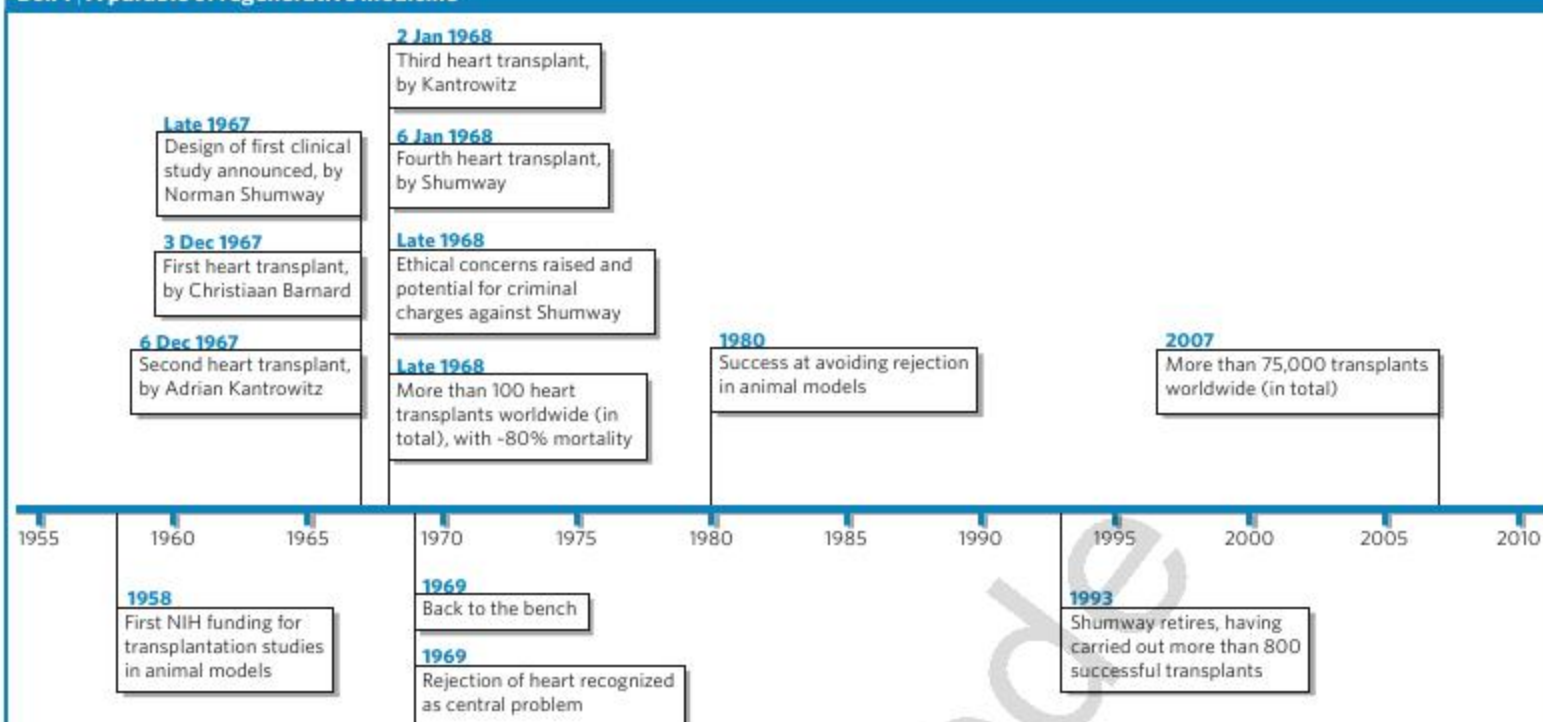
**Table 1 | Regenerative therapies for reversing advanced heart failure**

Therapy	Limitations	Status	Reference
Human heart transplantation	Not enough donors	Approved	See Box 1
Artificial heart implantation	Complications of blood clotting and continuous power source	Not yet approved for routine use	13
Ventricular-assist-device implantation	Complications of infection and stroke	Approved for short-term use	14
Cell-based therapy	Results mixed or marginal: optimal cell types and delivery system unclear, and low grafting efficiency	Not yet approved: in clinical trials	See page 322
Gene therapy to reverse calcium-cycling defects	Optimized heart-targeted in vivo delivery system unclear	Not yet approved: in early clinical trials	15

<sup>1</sup>Cardiovascular Research Center, Massachusetts General Hospital, Simches Research Center, 185 Cambridge Street, Boston, Massachusetts 02114, USA. <sup>2</sup>Department of Stem Cell and Regenerative Biology, Harvard University, 42 Church Street, Cambridge, Massachusetts 02138, USA.



## Box 1 | A parable of regenerative medicine



In the current era of molecular medicine, research that is beginning to unlock the potential of stem-cell biology has renewed the hope of patients and their families and has captured the imagination of a generation of doctors and scientists. But the road to modern regenerative medicine has been difficult, with more than its share of false starts, detours, dead ends and hazards. In searching for clues to the path ahead, it is instructive to look back to an earlier era.

One of the most important modalities of regenerative medicine today is heart transplantation. But, in the late 1950s, the idea of reversing heart failure by transplanting a heart from one individual to another was daring. The surgical techniques had not yet been established, and there were no clearly established approaches to preventing rejection of the new heart or preserving its function. The characteristics of a suitable donor heart had not yet been defined, and several legal and ethical dilemmas were waiting to be resolved. Undeterred, in 1958, Norman Shumway received his first grant for transplantation work from the National Institutes of Health (NIH), and he spent the next decade optimizing the surgical techniques for heart transplantation, mainly in animal models (see timeline). Towards the end of 1967, he announced that the work was ready for its first clinical application. But, shortly after (and before the regulatory issues concerning 'brain death' in the United States could be resolved), the first heart transplant was carried out in South Africa, by a young surgeon, named Christiaan Barnard, who had trained with Shumway. A media frenzy ensued and was described by Shumway as a "circus atmosphere with Marx Brothers overtones". Within the next year, more than 100 transplants were carried out around the world. However, the patients often survived only a few weeks, at most, after the operation, and by the end of 1968, the procedure was largely considered a failure and was abandoned.

Shumway's operative success in those early days was no different from

that of other surgeons, and he realized that the scientific underpinnings of the procedure needed to be identified to allow heart transplantation to move forward. So he returned to the laboratory, and during the next decade, he tackled the problem by using an interdisciplinary approach. His research team began to realize that the fundamental problem was not surgical technique but rejection of the heart, underscoring the need for stronger suppression of the immune system and improved monitoring of tissue rejection. The fundamental scientific advances made by Shumway and his team during the 1970s, together with the discovery of the immunosuppressant cyclosporine (almost 15 years after Shumway's initial funding from the NIH), paved the way for successful heart transplants.

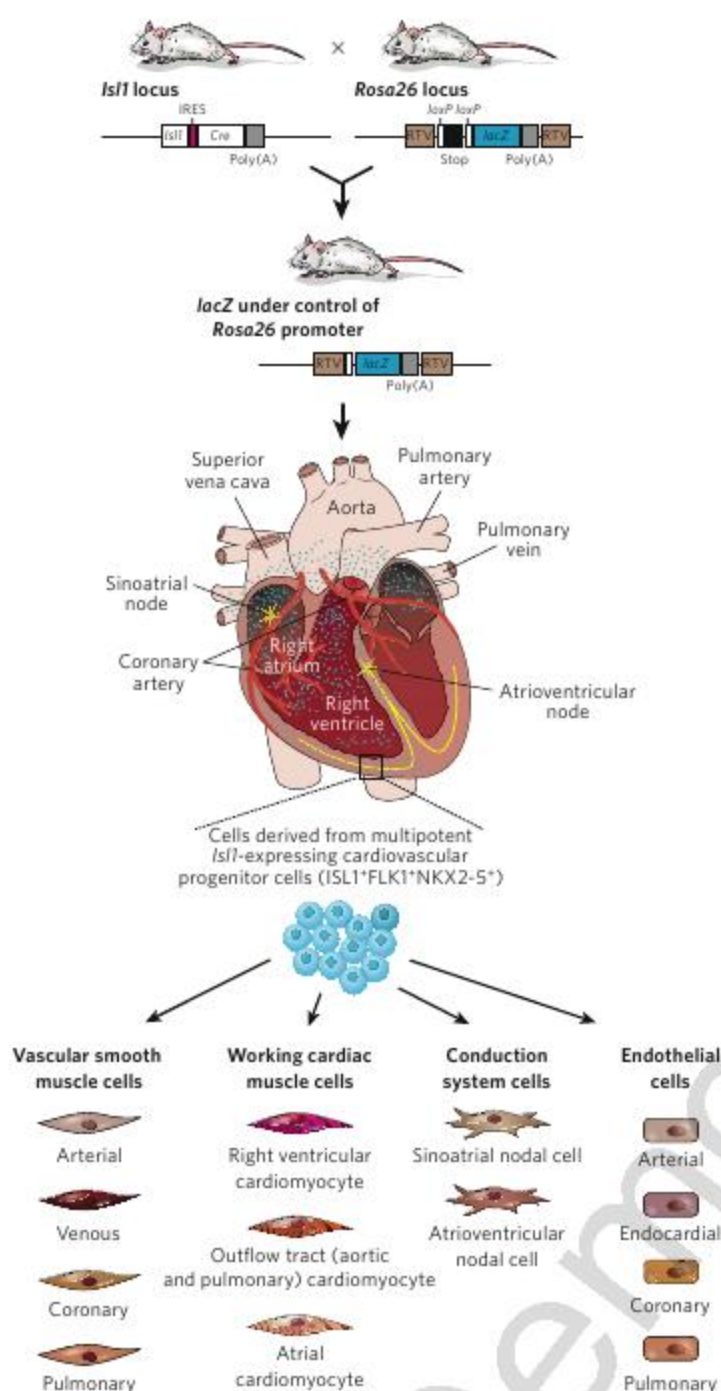
Shumway's work not only required the formation of interdisciplinary scientific and clinical teams with expertise in immunology, pathology, surgery, cardiology and transplantation biology, but also required the integration of advances in drug discovery driven by the pharmaceutical industry. During this work, he was careful not to raise false hopes, instead letting science drive the path to the clinic. At the press conference after his first transplant procedure, Shumway emphasized that too little was known about heart transplants in humans to promise an appreciable extension to a patient's life: "We have reached first base perhaps, but the work is just beginning." Controlling the immunological pathways of tissue rejection seems to be as crucial now for cell-based therapy (discussed by Ann Chidgey, Daniel Layton, Alan Trounson and Richard Boyd; see page 330) as it was in the early days of organ transplantation. Perhaps this same prescription — creating teams of doctors and scientists with a shared long-term vision, and involving the private sector — will be successful in the modern era of regenerative medicine. It might be just what Dr Shumway ordered.

reproducibly and then to fully differentiate enough of the cells of interest. In this regard, one of the main obstacles to establishing human ES-cell-based models is that ES cell lines vary. All lines do not have the same potential to differentiate into cells of a particular lineage, most probably as a result of inherent epigenetic, genetic and developmental differences at the time of their isolation. For example, a study of 17 independent human ES cell lines showed that 7 of these lines had little or no capacity to enter the cardiovascular lineage<sup>4</sup>, and the level of cardiovascular markers expressed by 2 of the 17 cell lines was an order of magnitude or more higher than that of these 7 lines. Similar variability between human ES cell lines was observed for entry to the pancreatic lineage, and cell lines that were optimal for generating cells

of endodermal lineages were extremely poor for generating mesodermal lineage cells in many cases<sup>4</sup>. Thus, new human ES cell lines that are optimal for generating specific lineages of interest need to be produced. In addition, iPS cell lines might be similarly variable. Even iPS cell lines derived from a single stock of patient-specific fibroblasts are likely to vary extensively, potentially making comparisons between disease-specific cell lines and normal cell lines challenging. It will therefore also be crucial to compare the proportion of highly purified, well-characterized differentiated cells of each lineage that can be obtained from various ES cell lines to set a baseline for variability.

The heterogeneity of progenitor cells that make up a particular organ presents an additional challenge. Most human ES cells can easily enter the





**Figure 1 | Multipotent cardiovascular progenitor cells and their differentiated progeny.** Lineage tracing during mouse heart development, by using irreversible genetic tagging with a *lacZ* reporter gene, shows that most of the cells in the mammalian heart, except those of the left ventricular myocardium, are derived from progenitor cells that express the gene islet 1 (*Isl1*; blue circles). More specifically, mice with the gene encoding the recombinase Cre under the control of the endogenous *Isl1* locus were crossed with mice from a LacZ indicator line. In the progeny, the stop sequence that prevents *lacZ* expression in the indicator line is deleted only in cells that express *Isl1*. Therefore, *lacZ* expression is irreversibly induced in these cells. By monitoring the distribution of the LacZ reporter in adult heart tissues, it is possible to trace the heart components that were derived from *Isl1*-expressing progenitor cells during the earliest stages of cardiogenesis. The differentiated progeny of multipotent *Isl1*-expressing progenitor cells contribute to all of the four main cell types in the heart (namely vascular smooth muscle cells, working cardiac muscle cells, conduction system cells and endothelial cells) in multiple cardiac compartments. Determining how these multipotent progenitor cells relate to other cardiovascular progenitor cells that have been reported, and identifying the precise pathways that direct their renewal and differentiation, is important for developing human models of human cardiovascular disease that are based on ES or iPS cells. FLK1, fetal liver kinase 1 (also known as KDR); IRES, internal ribosomal entry site; NKX2-5, NK2 transcription factor related, locus 5; RTV, retroviral integration sequence. (Figure adapted, with permission, from ref. 16.)

cardiovascular lineage and become beating cardiac muscle cells (cardiomyocytes). But, *in vivo*, multipotent cardiovascular progenitor cells can become multiple distinct cell types, including arterial and venous smooth muscle cells, conduction system cells that have endogenous pacemaker activity, endothelial cells of diverse lineages, and atrial and ventricular cardiomyocytes from various heart compartments<sup>5–8</sup> (Fig. 1). So, to set up an ES-cell-based model, it will be necessary to direct the differentiation of enough ES cells into a homogeneous population of specific progenitor cells for each cardiovascular cell type of interest. Thus, a more refined 'lineage fate map' is needed, to establish the relationship between these multipotent progenitor cells and other promising multipotent cardiovascular progenitor cells<sup>9,10</sup> and to identify the pathways that control their fate, directing them to become a specific type of mature adult cardiovascular cell (Fig. 1). As noted by Leonard Zon (see page 306), the identification of the self-renewal pathways for cellular intermediates that are committed to particular differentiated cardiovascular cell types, such as the completely committed cardiomyocyte progenitor, could prove extremely valuable for designing new strategies to regenerate cardiac muscle in the intact heart.

In patients, many crucial disease end points are best identified by studying the phenotype of intact organs. For the heart, the properties measured include electrical conduction, contractility, relaxation and repolarization. Many important properties are controlled by crosstalk between different tissues. Blood pressure, for example, is controlled by crosstalk between endothelial cells and vascular smooth muscle cells. Accordingly, when developing stem-cell-based disease models, it will be crucial to design physiological assay systems that can go beyond measurements in single cells and allow the measurements in intact tissues. One stem-cell-based model that could be valuable for such measurements is the *in vivo* grafting of purified human ES-cell-derived (or iPS-cell-derived) cardiovascular cell types into mouse models of cardiovascular disease, although several challenges need to be met before such cell transplantation is optimized (as reviewed by Robert Passier, Linda van Laake and Christine Mummery; see page 322). An alternative to transplantation is *ex vivo* tissue engineering. For example, thin films of synchronously contracting strips of cardiac muscle are being developed<sup>11</sup>, and these could be useful for measuring contractility and conduction. Progress is also being made towards modifying this technology so that specific heart parts can be engineered from multipotent cardiovascular progenitor cells. Recent studies suggest that the extracellular matrix has a crucial role in the assembly of differentiated cardiac cells into an intact heart<sup>12</sup>, and it is possible that each heart part has a 'signature' extracellular-matrix composition. Therefore, it is important to address directly how interstitial fibroblasts and other cells that secrete extracellular-matrix proteins are involved in controlling the fate of cardiovascular progenitor cells and stem cells (as discussed by Geoffrey Gurtner, Sabine Werner, Yann Barrandon and Michael Longaker; see page 314). Integrating technology from tissue engineering will be crucial for moving away from studies of single cells towards the analysis of physiological end points, opening the door both to new model systems and to the formation of heart tissue components *ex vivo* for direct transplantation.

### The potential benefits

The use of human stem-cell-based models will have a major impact on unravelling the molecular pathways involved in complex traits in humans. At a basic level, these systems could lead to a more complete understanding of human physiology at the cellular level. In terms of disease, these systems could allow the study, for example, of the complex common genetic variants that are now being uncovered by genome-wide association studies of patients with specific forms of heart disease, variants that would be extremely difficult to mimic in mice. Skin cells from adult patients with these variants could be used to generate iPS cells with ES-cell-like properties. In this manner, a new field of 'degenerative' medicine might be in the offing, allowing the discrimination of subsets of degenerative diseases, based on a combination of genetic variants that confer risk of developing a given disease phenotype. Subsequently, it should be possible to use patient-specific iPS cell lines to



study the functional effects of such genetic variants on differentiated cardiovascular cells.

This convergence of human genetics and human stem-cell models could be especially powerful for studying rare chromosomal disorders, such as Down's syndrome, that are usually associated with numerous physiological changes across multiple organ systems. The chromosomal translocations, deletions and duplications seen in many of these genetic disorders in which several organs are affected have been difficult, if not impossible, to study in model organisms. It could be particularly enlightening to examine the functional defects in differentiated cells derived from highly purified progenitor cell populations that harbour these chromosomal alterations.

The ability to isolate multipotent cells from specific patients and then obtain highly purified subsets of progenitor cells and their differentiated progeny is clearly important for the studies described so far, and it also provides an opportunity for identifying drug targets, screening drug candidates and measuring drug toxicity (as discussed by Yue Xu, Yan Shi and Sheng Ding; see page 338). In this case, it might be optimal to start with genetically engineered human ES cell lines that have a high inherent propensity to enter the particular lineage of interest. 'Knock-ins' of mutant genes that correspond to monogenic diseases, as well as genetic reporters that allow progenitor cells and their differentiated progeny to be identified and purified by flow cytometry, will be extremely valuable for such assays. Gene targeting in human ES cell lines is, however, notoriously inefficient, but new tools for this technique will be available soon. An alternative strategy is to use iPS cell lines derived from patients with monogenic diseases. Because there are hundreds of differences between the transcriptional profiles of iPS cell lines and authentic ES cells, it will be crucial to ensure that the differentiated progeny of iPS cell lines have fidelity to the fully differentiated (mature) human cell types of interest (for example, iPS-cell-derived cardiomyocytes will need to have the same physiological phenotype as endogenous cardiomyocytes under baseline conditions and during extended periods of mechanical work). If this can be accomplished, then human-cell-based models of human heart disease should offer a new and direct means of identifying chemical compounds, secreted factors and antibodies that can ameliorate the disease phenotype.

### Beyond stem-cell biology

If humans are the model organism, then cross-disciplinary collaborations between scientists and doctors will be essential. Advances in understanding the genetics of complex traits in humans, in the imaging of cells *in situ* in the human body (as discussed by Timm Schroeder; see page 345) and in the engineering of human tissue will be necessary to move beyond stem-cell biology towards regenerative medicine. One aspect that holds considerable promise is the possibility of identifying groups of patients who correspond to susceptibility or resistance to

disease phenotypes of interest and for whom the relationships between genotype and phenotype can be assessed both *in vivo* and *in vitro* in stem-cell-based model systems. 'Calling back' individuals of interest for further *in vivo* physiological study is likely to be a crucial component of these model systems, bringing about a shift in the paradigm of separate clinical and scientific studies. For these studies to succeed, it will be imperative for researchers to forge close ties with major academic hospitals that have access not only to the phenotyping tools and the appropriate patients but also to doctors with extensive training in stem-cell biology, who are likely to drive the field in the coming years. Perhaps we are about to enter a new age of enlightenment in regenerative medicine, with human models of human disease leading the transition from bedside to bench.

1. Cowan, C. A. *et al.* Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350**, 1353-1356 (2004).
2. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920 (2007).
3. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).
4. Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nature Biotechnol.* **26**, 313-315 (2008).
5. Laugwitz, K. L., Moretti, A., Caron, L., Nakano, A. & Chien, K. R. Islet1 cardiovascular progenitors: a single source for heart lineages? *Development* **135**, 193-205 (2008).
6. Laugwitz, K. L. *et al.* Postnatal Isl1<sup>+</sup> cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433**, 647-653 (2005).
7. Moretti, A. *et al.* Multipotent embryonic Isl1<sup>+</sup> progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151-1165 (2006).
8. Qyang, Y. *et al.* The *in vivo* renewal and differentiation of Isl1<sup>+</sup> cardiovascular progenitors is controlled by a Wnt/ $\beta$ -catenin pathway. *Cell Stem Cell* **1**, 165-179 (2007).
9. Wu, S. M. *et al.* Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127**, 1137-1150 (2006).
10. Kattman, S. J., Huber, T. L. & Keller, G. M. Multipotent flk-1<sup>+</sup> cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev. Cell* **11**, 723-732 (2006).
11. Feinberg, A. W. *et al.* Muscular thin films for building actuators and powering devices. *Science* **317**, 1366-1370 (2007).
12. Ott, H. C. *et al.* Perfusion-decellularized matrix: using nature's platform to engineer a bioartificial heart. *Nature Med.* **14**, 213-221 (2008).
13. Slepian, M. J. & Copeland, J. G. The total artificial heart in refractory cardiogenic shock: saving the patient versus saving the heart. *Nature Clin. Pract. Cardiovasc. Med.* **5**, 64-65 (2008).
14. Mather, P. J. & Konstam, M. A. Newer mechanical devices in the management of acute heart failure. *Heart Fail. Rev.* **12**, 167-172 (2007).
15. Kaye, D. M., Hoshijima, M. & Chien, K. R. Reversing advanced heart failure by targeting Ca<sup>2+</sup> cycling. *Annu. Rev. Med.* **59**, 13-28 (2008).
16. Martin-Puig, S., Wang, Z. & Chien, K. R. Lives of a heart cell: tracing the origins of cardiac progenitors. *Cell Stem Cell* **2**, 320-331 (2008).

**Acknowledgements** The author is supported by grants from Massachusetts General Hospital, the Harvard Stem Cell Institute, Fondation Leducq and the National Institutes of Health. This article is dedicated to N. E. Shumway.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The author declares competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to the author ([kcchien@partners.org](mailto:kcchien@partners.org)).



# Intrinsic and extrinsic control of haematopoietic stem-cell self-renewal

Leonard I. Zon<sup>1</sup>

**When stem cells divide, they can generate progeny with the same developmental potential as the original cell, a process referred to as self-renewal. Self-renewal is driven intrinsically by gene expression in a cell-type-specific manner and is modulated through interactions with extrinsic cues from the environment, such as growth factors. However, despite the prevalence of the term self-renewal in the scientific literature, this process has not been defined at the molecular level. Haematopoietic stem cells are an excellent model for the study of self-renewal because they can be isolated prospectively, manipulated relatively easily and assessed by using well-defined assays. Establishing the principles of self-renewal in haematopoietic stem cells will lead to insights into the mechanisms of self-renewal in other tissues.**

Of all the known types of stem cell, haematopoietic stem cells (HSCs) are the best characterized<sup>1</sup>. HSCs were first proven to exist in 1961, and since then a wealth of experimentation has formed a solid foundation for the study of stem cells. HSCs can differentiate to form mature blood cells but can also reproduce themselves, which is known as self-renewal. Self-renewal is a biological process in which a cell with long-term differentiation capability generates progeny with the same potential to differentiate into various cell types. HSCs reside in distinct stem-cell niches that vary in location depending on the developmental stage of the organism. The formation of blood cells begins in the yolk sac of the embryo and then shifts to the aortic region, and then sequentially to the placenta, fetal liver and adult bone marrow. Except for the yolk sac, in which mainly erythrocytes form, the subsequent sites contain HSCs with the ability to self-renew.

When studying the haematopoietic system, self-renewal is assessed by using a transplantation assay. Transplantation of a single HSC from the bone marrow into a lethally irradiated animal can rescue the entire immune system. HSCs can then be isolated using a cocktail of monoclonal antibodies specific for various cell-surface proteins; for example, expression of the markers Kit and stem-cell antigen 1 (SCA1) but not lineage (Lin) markers defines an HSC-containing cell population (known as KSL cells). In this population of cells, one in five cells has the property of long-term engraftment, as defined in transplantation assays. In secondary assays, HSCs are isolated from an animal that received an HSC transplant and then transplanted into a different recipient, establishing the extensive self-renewal capacity of HSCs. Recent studies with other monoclonal antibodies such as CD150-specific antibody have led to further enrichment of HSCs<sup>2</sup>.

Self-renewal occurs in a cell-autonomous (intrinsic) manner but is influenced by the environment. Stem-cell niches are important regulators of self-renewal that affect whether symmetrical or asymmetrical stem-cell division occurs. During symmetrical division, the 'parent' stem cell forms either two parent stem cells or two daughter cells, which are differentiated. By contrast, during asymmetrical division, the 'parent' stem cell forms another parent stem cell and a daughter cell. Asymmetrical division has been described to occur in several tissue-specific stem-cell populations. For example, a neural stem cell can remain a neural

stem cell or can differentiate into neurons<sup>3</sup>. In the fruitfly, *Drosophila melanogaster*, the male germline stem cell resides in an environment called the niche, attached to a hub cell, and divides to become another germline stem cell or differentiates into a sperm cell<sup>4</sup>. And, in some cancers, the transformation of a progenitor cell can lead to activation of self-renewal, a property of 'cancer stem cells'<sup>5</sup>. In each of these cases, the molecular steps of self-renewal have yet to be defined. In terms of HSCs, the functional heterogeneity that is observed in HSC populations in culture has been linked to asymmetrical division<sup>6</sup>, and the probability that asymmetrical division will occur is influenced by the presence of certain cytokines<sup>7</sup>. Furthermore, some proteins have been shown to be distributed asymmetrically during HSC division<sup>8</sup>. For example, recent studies in *D. melanogaster* showed that the parent centrosome is inherited asymmetrically during stem-cell division<sup>9</sup>.

In this review, I delineate the characteristics of HSC self-renewal, as well as the molecular pathways that are known to be involved in the self-renewal process. It is probable that self-renewal is regulated in a distinct manner in each tissue, with different molecular pathways being involved, but there might be some common control processes.

## Self-renewal of HSCs and stem cells in other organs

Because a single HSC can self-renew and differentiate into cells of the haematopoietic lineages, this cell can, in essence, function for the lifetime of the animal. Apart from blood, several other tissues are known to self-renew. *D. melanogaster* testes contain germline stem cells that produce sperm for most of the animal's life<sup>4</sup>, and the transplantation of testis tissue from one mammal into another indicates the presence of stem-cell activity<sup>10</sup>. In addition, skin grafts last for a lifetime, indicating that stem cells reside in the skin<sup>11</sup>. Mammary cells transplanted into the mammary fat pad also provide long-term reconstitution of breast tissue<sup>12</sup>. Furthermore, gut stem cells continuously yield the cells of an entire villus, indicating that a self-renewal mechanism exists in the gut<sup>13</sup>. And, neural stem cells can be cultured to form neurospheres that can be replated to form new neurospheres and transplanted into the brain, where they contribute to neuronal function, as determined by measuring synapse formation and electrical response. Finally, muscle satellite cells, a population of adult muscle stem cells, can be isolated

<sup>1</sup>Stem Cell Program and Division of Hematology/Oncology, Children's Hospital Boston, Karp 7, 1 Blackfan Circle, Boston, Massachusetts 02115, USA.



and injected into muscle, thereby contributing to muscle function for considerable time periods<sup>14</sup>. Despite these similarities, the process of self-renewal might differ in each of these tissues, and some adult tissues, such as the pancreas, do not contain any stem cells<sup>15</sup>.

Embryonic stem (ES) cells also have a self-renewal programme<sup>16</sup>. ES cells can be kept in an undifferentiated state *in vitro* by contact with feeder cells or, in the case of mouse ES cells, by adding antidifferentiation factors such as leukaemia inhibitory factor to the culture media. In these conditions, the cells divide without differentiating, thereby renewing themselves. These undifferentiated ES cells can be transferred to another culture plate with similar antidifferentiation conditions, and they will continue to renew themselves. At any stage in this process, the cells can be transferred to a culture medium that promotes differentiation, and maturation into tissue cells will then occur. When the undifferentiated cells are injected into mice, teratomas (tumours that contain heterogeneous tissues) develop, suggesting that ES cells can differentiate *in vivo*. Undifferentiated ES cells can also generate all of the tissues of the body when injected into a blastocyst. Therefore, ES cells can self-renew indefinitely in culture, while maintaining the ability to differentiate.

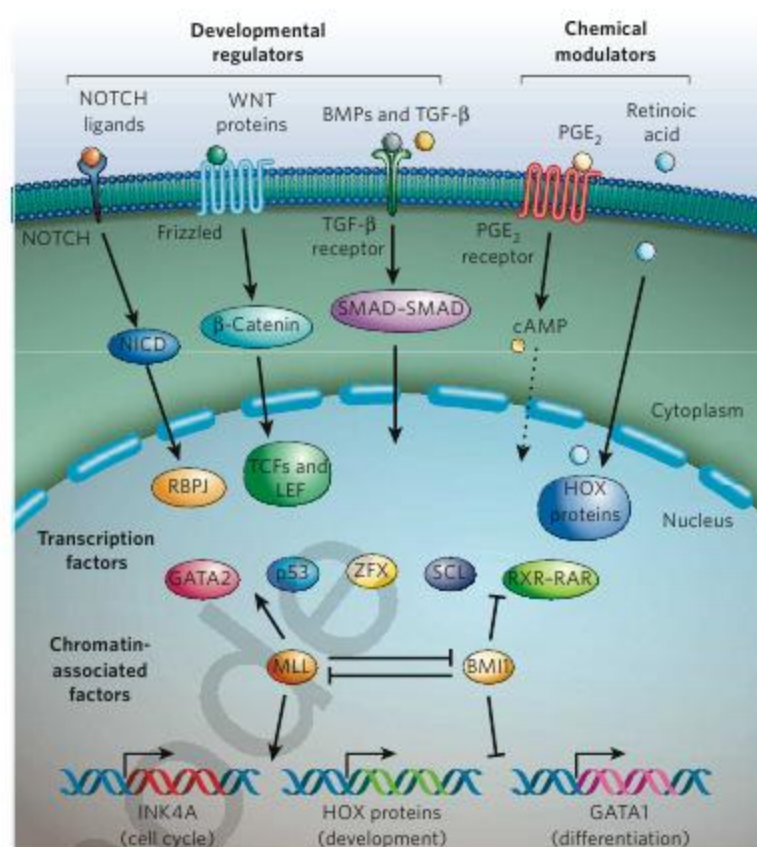
### A hierarchy of stem cells and progenitor cells

In most tissues, stem cells form progenitor cells with a more limited differentiation potential. On the basis of transplantation experiments with mouse haematopoietic cells, transplanted HSCs can be classified into two distinct populations: short-term engrafting and long-term engrafting. Long-term-engrafting HSCs can produce blood cells for at least 6 months, and these HSCs can be transferred to new hosts by secondary transplantation. By contrast, short-term-engrafting HSCs contribute to the peripheral blood for up to 3 months after transplantation, possibly making a skewed contribution to the myeloid or lymphoid cell compartments. Most progenitor cells and further-differentiated cells cannot self-renew, although certain types of differentiated cell can (for example, lymphocytes).

In addition to such transplantation assays, the developmental hierarchy of stem cells and progenitor cells can be revealed by examining the cell cycle. Cells in the skin that take up 5-bromodeoxyuridine (BrdU), a marker of cycling cells, have been called transient amplifying cells and are thought to be early progenitor cells that are entering the cell cycle. Recent evidence suggests that, in some cases, transient amplifying cells in an organ can replace exhausted stem cells<sup>17</sup>. In this study, after the animals had been treated with pulses of BrdU, the testes were examined by using a DNA-recombination-based technique to track the fate of stem cells and their immediate progeny. At 3 months after pulse BrdU labelling, 0.3% of normal testis cells were labelled. After transplantation of these labelled testis cells into testes that had been genetically engineered to lack stem cells, about 12% of the germline cells were labelled, indicating that the transient amplifying cells were contributing to the stem-cell population in the transplant setting. The contribution of transient amplifying cells was found to be similar after chemotherapy had been used to deplete the stem cells. These transient amplifying cells in testes seem to acquire self-renewal potential when they incorporate into the stem-cell niche<sup>17</sup>. In other tissues, such as the skin, transient amplifying cells leave the stem-cell niche and contribute to the differentiated tissue<sup>18</sup>.

Stem cells from a single tissue but at different developmental stages, such as fetal HSCs and adult HSCs, can show distinct differences in self-renewal properties. A new developmental checkpoint that marks development into an adult HSC has been defined by investigating serially cultured HSCs and carrying out transplantation experiments<sup>19</sup>. In this study, the fetal HSCs were shown to have an intrinsic ability to self-renew and differentiate, and this property was maintained until 1 week after birth. At this time, the HSCs switched to being quiescent, similar to adult HSCs. The switch was subsequently shown to require the production of stage-specific factors, such as the transcription factor SOX17, in fetal HSCs and not adult HSCs<sup>20</sup>.

It is unclear whether the development of organs in the fetus and embryo requires self-renewal. It is possible that mouse HSCs arise in the



**Figure 1 | Developmental signalling pathways involved in HSC self-renewal.** When developmental regulators and chemical modulators bind to the appropriate cell-surface receptors, signalling pathways are initiated, leading to the translocation of transcription factors from the cytoplasm to the nucleus. (The receptor for retinoic acid is in the nucleus.) These transcription factors then interact with other cell-specific transcription factors to regulate self-renewal. Competition between transcription factors provides a simple mechanism for controlling self-renewal, and chromatin-associated factors, such as MLL and BMI1, provide another level of control (and competition). BMP, bone morphogenetic protein; cAMP, cyclic AMP; LEF, lymphoid-enhancer-binding factor; NICD, NOTCH intracellular domain; PGE<sub>2</sub>, prostaglandin E<sub>2</sub>; RXR, retinoid X receptor; TCF, T-cell factor; TGF-β, transforming growth factor-β; ZFX, X-linked zinc-finger protein.

aorta–gonads–mesonephros region and the placenta before embryonic day 12.5 and that all of the HSCs required for both fetal and adult life are amplified from these original, developmentally specified cells<sup>21</sup>. This idea implies that a cell's capacity for self-renewal is conferred during early development but is also used later, for the development of mature organs.

### Self-renewal is molecularly linked to developmental pathways

Tissue development and differentiation are regulated by several intracellular signalling pathways. These pathways are activated by growth factors and morphogens such as NOTCH ligands, WNT proteins and bone morphogenetic proteins (BMPs) (Fig. 1), triggering spatially overlapping signals that are involved in the development of almost all organs that have been studied in detail. As shown in Fig. 1, when a ligand binds to its receptor, this results in the initiation of signalling pathways that culminate in the nucleus. Some ligands are produced only in particular tissues, although these spatially distinct ligands might activate common downstream signalling mechanisms. For example, WNT2 is crucial for the generation of a liver<sup>22</sup>, whereas WNT1 is required for the production of specific types of neuron in the brain<sup>23</sup>. Signalling mediated by either of these factors activates β-catenin, presumably triggering the transcription of target genes encoding the cues that are required to initiate and specify the development of the respective tissues.

In later phases of fetal and adult development, these pathways have marked effects on the self-renewal process, leading to an increase



in the size of stem-cell pools (Table 1). Addition of WNT3A to the culture medium in which specific bone-marrow subpopulations are growing results in a threefold increase in self-renewal, as measured by the number of HSCs that engraft after transplantation into irradiated recipients<sup>24</sup>. In addition, the growth factor sonic hedgehog (SHH) can increase the self-renewal of adult neural progenitor cells<sup>25</sup>. Therefore, as well as contributing to the early designation of

cell types, these developmental regulators can act directly on more mature stem-cell populations.

For each of the pathways mentioned, receptor–ligand interactions result in transcriptional activation. WNT proteins, hedgehog proteins, BMPs and the NOTCH ligand known as Delta-like stimulate the activation and translocation of transcription factors into the nucleus, where they activate or repress the genes involved in organ development. For the NOTCH pathway activated by Delta-like, the intracellular domain of NOTCH is cleaved and translocates to the nucleus, where it interacts with RBPJ (a homologue of the *D. melanogaster* protein Su(H)) to regulate target genes. By contrast, WNT proteins function to stabilize  $\beta$ -catenin, which travels to the nucleus and interacts with lymphoid-enhancer-binding factor (LEF) and members of the T-cell factor (TCF) family, driving expression of genes that encode proteins such as Myc and cyclin D. Hedgehog-mediated signalling leads to the expression of genes encoding GLI-family factors, which are transcriptional regulators, and BMPs activate SMADs, which translocate to the nucleus after activation. Because growth-factor-mediated signalling leads to a quick transcriptional response, it is an ideal mechanism to drive organogenesis, as well as to modulate self-renewal later in development.

### Pathways for stem-cell homeostasis versus tissue repair

Stem cells must be maintained so that there is a continuous supply for steady-state organ function. For example, in *D. melanogaster* with a mutation in the gene encoding an orthologue of the transcription factor JAK, the stem cells in the gonads arise and then differentiate only once<sup>26</sup>. This loss of self-renewal leads to lack of fertility. Another example is that mice that lack the gene *Bmi1*, which encodes a chromatin-associated factor, are born with stem cells, but these cells seem to exhaust themselves: the animals progressively lose blood and neural cells<sup>27</sup>. In addition to their role in these homeostatic processes, stem cells are also necessary after tissue damage has occurred or during stress or regeneration. The pathways that activate stem cells in these cases might be distinct from those that regulate stem-cell homeostasis.

A list of the factors that are necessary and/or sufficient for self-renewal is provided in Table 1. How some of these factors function is controversial. For example, when an inducible Cre recombinase construct is used to generate mice with a conditional knockout of the gene encoding  $\beta$ -catenin (as well both  $\beta$ -catenin and  $\gamma$ -catenin) in the bone marrow, the mice have normal numbers of haematopoietic cells and normal numbers of HSCs, despite the block in the WNT-mediated signalling pathway<sup>28</sup>. By contrast, when a haematopoietic-cell-specific Cre construct is used to make a conditional knockout of the gene encoding  $\beta$ -catenin in haematopoietic tissue, although the outcome would be expected to be similar, the mutant mice have defects in self-renewal during fetal and adult HSC homeostasis<sup>29</sup>. Nevertheless, adding WNT3A to the culture medium is sufficient to increase self-renewal in those mice<sup>24</sup>. The role of Wnt proteins in tail-fin regeneration in zebrafish (*Danio rerio*)<sup>30</sup> is illustrated in Fig. 2. After the tail has been removed, a unique tissue forms. This tissue, called a blastema, has the ability to self-renew and ultimately differentiates to reform the tail in a wild-type fish. If *wnt8* is overexpressed, proliferation increases, whereas overexpression of another member of the *wnt* gene family, *wnt5*, suppresses the recovery of the tail fin. Similarly, in other tissues, activation of the *wnt3* and *wnt8* family (the canonical Wnt pathway) increases the recovery of cells after damage, and overexpression of *wnt5* stimulates a pathway that suppresses tissue recovery that does not include  $\beta$ -catenin.

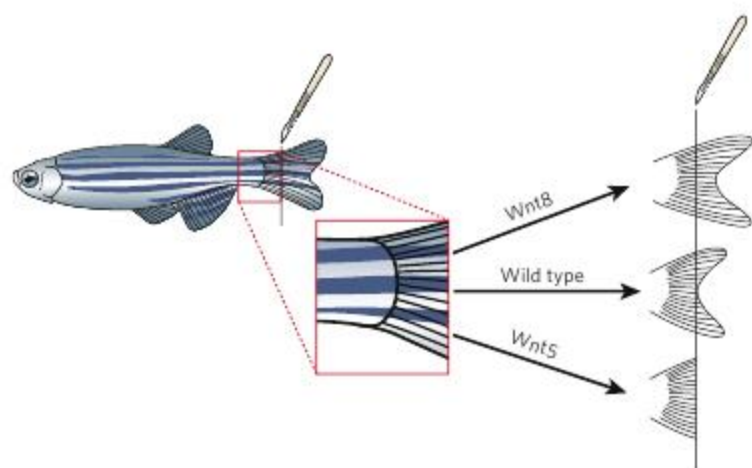
SMAD4, a component of the BMP- and transforming growth factor- $\beta$  (TGF- $\beta$ )-mediated signalling pathways, is required for normal HSCs to self-renew following transplantation<sup>31</sup>. For other tissues, such as the adult central nervous system and hair follicles, BMP- and TGF- $\beta$ -mediated signalling pathways are required for homeostasis of the tissue stem cells and progenitor cells. Another type of growth factor, fibroblast growth factor 20 (Fgf20), seems to be specifically used during regeneration of zebrafish tail fins but is not required for normal growth<sup>32</sup>. These growth-factor-mediated signalling pathways might be constantly stimulated in small populations of cells in each organ, thus modulating

**Table 1 | Developmental regulators and self-renewal factors for HSCs**

Factor (and transcription-factor output)*	Function in self-renewal process	References‡
<b>Growth factor or chemical modulator</b>		
NOTCH ligands (RBPJ and NICD)	Not necessary, sufficient	73, 74
WNT proteins ( $\beta$ -catenin, TCFs and LEF)	Might be necessary†, sufficient	24, 28
BMPs (SMADs)	Maintains stem cells; SMAD4 necessary	31
Angiopoietin-like factors	Sufficient	75
Thrombopoietin	Necessary	76
Kit ligand	Necessary	33
Retinoic acid (HOX proteins)	Necessary, sufficient	60
PGE <sub>2</sub>	Might be necessary†, sufficient	56
<b>Cell-cycle regulators</b>		
INK4A (also known as p16)	Increases (in older animals)	77
INK4C (also known as p18)	Increases	78
KIP1 (also known as p27)	Increases (in <i>Mad1</i> mutants)	79
WAF1 (also known as p21)	Decreases (on mixed but not pure background)	80
PTEN	Decreases	81, 82
p53	Decreases	83
Myc	Decreases	66
<b>CDX and HOX proteins</b>		
Hoxa7	Increases (when overexpressed with cofactors)	84
Hoxa9	Increases (when overexpressed)	40
Hoxa10	Increases (when overexpressed)	41
Hoxb4	Increases (when overexpressed)	39
CDX2	Overexpressed in leukaemia	85
CDX4	Overexpressed in leukaemia	85
MEIS1	Sufficient (with <i>Hox</i> genes)	86
<b>Blood transcription factors</b>		
GATA2	Necessary	64
GFI1	Necessary	65
TEL	Necessary for HSC survival	87
JUNB	Necessary	88
SOX17	Necessary in fetus	20
PU.1	Necessary	89
Myb	Necessary	90
CBP	Necessary	91
ZFX	Necessary	70
RUNX1	Not necessary	92
SCL	Not necessary	93
<b>Chromatin-associated factors</b>		
EZH2	Necessary	94
BMI1	Necessary	44, 45
MLL	Necessary	54
RAE28	Necessary	95

BMP, bone morphogenetic protein; CBP, CREB (cyclic-AMP-responsive-element-binding protein)-binding protein; CDX, caudal-type homeobox; EZH2, enhancer of zeste homologue 2; LEF, lymphoid-enhancer-binding factor; NICD, NOTCH intracellular domain; PGE<sub>2</sub>, prostaglandin E<sub>2</sub>; TCF, T-cell factor; ZFX, X-linked zinc-finger protein. \*For growth factors and chemical modulators, the transcription factor output is listed if known. †Might be necessary indicates that this requirement has not been formally studied. ‡References are representative but limited in number because of space constraints.





**Figure 2 | The Wnt-mediated signalling pathway modifies tissue regeneration.** Zebrafish tail-fin regeneration is an excellent model for studying regeneration, and several factors that modulate regeneration have been found in this system. After the distal tail fin has been removed, a cell population called a blastema, which has the capacity to self-renew and differentiate, forms near the wound. The fin typically regenerates to a normal wild-type size. However, if the canonical Wnt-mediated signalling pathway is activated by Wnt8, cell proliferation increases in the regenerating fin (although the size of the regenerating fin eventually becomes equivalent to that of a wild-type fin). By contrast, activation of the non-canonical Wnt-mediated signalling pathway by Wnt5 suppresses regeneration.

cell behaviour, but they are not required for the overall function of the organ in the absence of regeneration or injury.

Specific signalling pathways can have different effects at different stages of stem-cell development. For example, in haematopoietic cells, signalling mediated by the growth factors Kit ligand and thrombopoietin can differentially regulate the responsiveness of a fetal HSC and an adult HSC<sup>33</sup>. Both fetal and adult HSCs were found to be Kit dependent *in vivo* and *in vitro*, although the fetal cells that cycled faster and were more efficient at dividing symmetrically were stimulated at a lower concentration of Kit ligand than adult cells. Fetal HSCs that were Kit-deficient cycled similarly to normal adult HSCs, suggesting that there is an intrinsic link between the response of the cell cycle and growth-factor stimulation in HSCs.

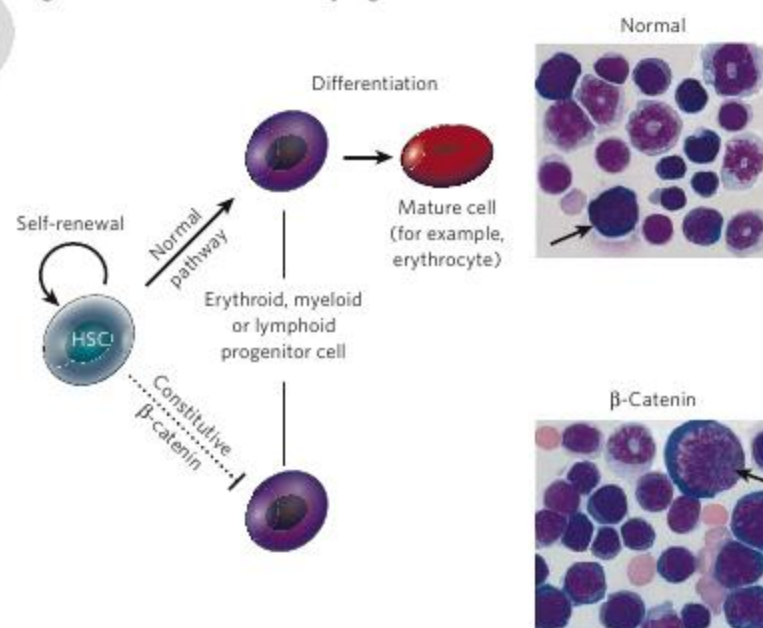
Continuous reactivation of the self-renewal programme can lead to stem-cell exhaustion. For example, when a gene encoding a stabilized form of  $\beta$ -catenin was expressed in mouse HSCs *in vivo*, the mice developed aplastic anaemia, as shown by a lack of HSCs and cells of the erythroid lineage<sup>34,35</sup>. As shown in Fig. 3, the constitutively activated  $\beta$ -catenin leads to a prominent defect in haematopoiesis, with differentiation being blocked at the erythroblast stage. Given the importance of maintaining a balance between HSC renewal and differentiation, it is possible that pulse activation by developmental regulators (such as transient treatment of an HSC population with the ligand WNT3, which is upstream of  $\beta$ -catenin) might lead to increased self-renewal, but chronic activation will deplete HSCs or force them into quiescence.

### The CDX-HOX pathway

The expression of the *HOX* gene family is intricately linked to the activation of self-renewal in HSCs; however, the precise role of *HOX* genes in the self-renewal process remains to be determined. Classical genetics experiments in *D. melanogaster* and mice have shown that *HOX* proteins function during the development of the body plan. Disruption of genes encoding specific *HOX*-family members in the four paralogous mammalian gene clusters leads to abnormalities in specific segments of the body. *HOX* gene expression is regulated by several mechanisms. The gene products of the caudal-type homeobox family (CDX1, CDX2 and CDX4 in vertebrates) bind directly to enhancers within *HOX* genes, thereby activating the genes. Other genes such as those encoding members of the PBX and MEIS families (which are *HOX* cofactors) also modulate *HOX* gene expression during development. In addition,

loss-of-function and gain-of-function studies have shown that the BMP, NOTCH-ligand- and hedgehog-mediated signalling pathways also modulate *HOX* gene expression and function during embryogenesis. For example, production of WNT proteins leads to aberrant *HOX* gene expression<sup>36</sup>, and WNT-mediated signalling is known to directly stimulate *Cdx* genes in mice<sup>37</sup>.

The link between self-renewal and *HOX* gene expression is most evident in the haematopoietic system. Zebrafish with a mutant *cdx4* lack HSCs during development<sup>38</sup>. CDX proteins function to induce expression of *HOX* genes, and it was found that injection of *hoxa9* messenger RNA rescues haematopoietic development in the *cdx4* mutant<sup>38</sup>. Overexpression of *Hoxb4* or *Hoxa9* in mouse bone-marrow cells leads to increased self-renewal of HSCs<sup>39,40</sup>. As shown in Fig. 4, mouse haematopoietic cells can be cultured *in vitro* and then adoptively transferred to mice, and differentiation proceeds. By contrast, overexpression of *Hoxb4* or *Hoxa9* leads to an increase in myeloblast and erythroblast formation *in vitro*, which is associated with improved self-renewal. However, after adoptive transfer, the cells can differentiate. The function of these two factors seems to be distinct: *Hoxa9* expression leads to increased myeloid lineage differentiation and leukaemogenesis, whereas *Hoxb4* expression leads to relatively normal bone-marrow development. Recently, in experiments in which moderate amounts of HOXA10 were introduced into HSCs, self-renewal was found to increase 15-fold, but high doses of HOXA10 blocked megakaryocytic and erythroid cell fates<sup>41</sup>. Factors that regulate *HOX* gene expression and function have also been found to participate in leukaemia. These include *PBX* and *CDX* genes, the retroviral insertion gene *MEIS1* and the partners of leukaemic fusion proteins. Similarly, the expression of *Pbx*, *Cdx* or *Meis* genes in mouse haematopoietic cells leads to increased self-renewal<sup>42</sup>. Taken together, the studies described in this section strongly suggest that *HOX* gene expression is important in the regulation of self-renewal, although loss-of-function experiments involving the *Hox* cluster genes in mice have only recently revealed defects in self-renewal<sup>43</sup>. It will be important to determine the nature of *HOX* targets in adult stem cells and progenitor cells.



**Figure 3 | Pathways that are involved in HSC homeostasis or in tissue regeneration can exhaust HSCs if activated frequently.** A simple schematic of the pathways of HSC self-renewal and differentiation is shown, together with micrographs of stained bone-marrow cells. Arrows indicate proerythroblasts, which differentiate into erythroblasts and then mature erythrocytes. HSCs can yield progenitor cells that usually differentiate into mature erythroid, myeloid and lymphoid cells (normal pathway). Introduction of a gene encoding a constitutively active  $\beta$ -catenin, which functions downstream of WNT proteins, into mouse HSCs results in a partial block (dashed blocking arrow) in differentiation, with erythroblasts evident in large numbers. The mice then become anaemic as a result of the block in the differentiation pathway, as well as HSC exhaustion caused by increased cycling and self-renewal. (Images reproduced, with permission, from ref. 34.)



### Epigenetic regulation of self-renewal

Certain transcription factors, such as HOX proteins, function together with chromatin-associated factors. Chromatin-associated factors, in turn, can modulate the expression or the function of specific DNA-binding proteins and coactivators involved in self-renewal. Such factors that stimulate or repress the modification of histones (and therefore regulate gene expression by epigenetic mechanisms) can have profound effects on cell fate and stem-cell function. Probably the best example of a chromatin-associated factor involved in self-renewal is BMI1 (Fig. 1 and Table 1), which is encoded by a polycomb group gene<sup>27,44–46</sup>. The Polycomb complex was first described in *D. melanogaster* as a series of genes that results in abnormal *Hox* gene expression when mutated; the gene products of the complex antagonize factors that are required for the maintenance of *Hox* gene expression<sup>44,45</sup>. For example, dominant alleles at the Polycomb locus of *D. melanogaster* are associated with homoeotic transformations of the thorax to the legs, and the antennae to the legs, effects that are caused by aberrant *Hox* gene expression.

In vertebrates, polycomb group proteins participate mainly in two complexes, polycomb repressive complex 1 (PRC1) and PRC2. BMI1 is a component of PRC1, which functions as a transcriptional repressor. Mice with homozygous null mutations in *Bmi1* (*Bmi1*<sup>−/−</sup> mice) are born with HSCs and neural progenitor cells that become exhausted, indicating that these cells have a decreased capacity for self-renewal; conversely, overexpression of *Bmi1* increases self-renewal<sup>47</sup>. This seems to be the case for stem cells in many tissues, including the blood, the nervous system, the breasts and the gastrointestinal tract<sup>44,45,48–52</sup>. BMI1-deficient mice produce large amounts of the cell-cycle regulator INK4A (also known as p16), indicating that BMI1 usually represses the gene encoding this regulator<sup>48,51</sup>. In support of this conclusion, the self-renewal defect in these mice can be partly rescued by mating them with mice that lack INK4A and/or another cell-cycle regulator, ARF (also known as p19)<sup>48,51</sup>. Taken together, these findings point to a direct mechanism by which a chromatin-associated factor can modulate the cell cycle and self-renewal.

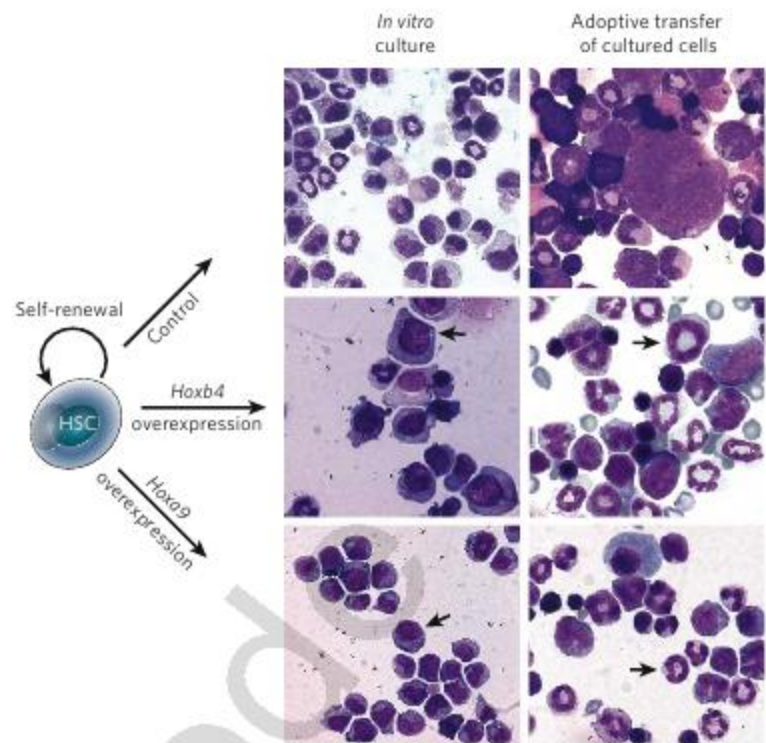
There is also another example of epigenetic control by the polycomb complexes in vertebrates. The gene *MLL* is an orthologue of *trithorax*, a gene regulated by the Polycomb complex in *D. melanogaster* (Fig. 1 and Table 1); *MLL* is often fused to another protein in leukaemic fusion proteins. The *MLL* fusion proteins in HSCs lead to increased self-renewal<sup>53</sup>, and inactivation of *MLL* in mice leads to decreased HSC self-renewal<sup>54</sup>.

Methylation or demethylation of crucial genes is another possible mechanism for the modulation of self-renewal. Recently, the DNA methyltransferases DNMT3A and DNMT3B were linked to self-renewal of HSCs<sup>55</sup>. The function of these two factors during haematopoiesis was studied by using a competitive haematopoietic transplantation system, in which the genes encoding these factors were excised only in HSCs. When each gene was deleted, there was no effect on self-renewal, but HSCs in which both genes had been deleted failed to compete with normal HSCs in repopulation assays. These experiments suggest that both genes are required for DNA methylation in HSCs and that this methylation has an important role in self-renewal.

The studies described in this section establish that epigenetic alterations can modulate the self-renewal process. Epigenetic changes in stem cells are not permanent and can be erased (partly or completely) by cell division. Therefore, such changes might facilitate the transition of a progenitor cell to a self-renewing stem cell, or might prompt a stem cell to differentiate, divide or lose the ability to self-renew.

### Chemical modulators of self-renewal

Several research groups have carried out chemical screening for compounds that modulate self-renewal in a variety of tissues and organisms (see page 338). The zebrafish is an excellent model system for such screens, facilitating, for example, screens for chemicals that modulate stem-cell number in the embryonic aorta (which need to be carried out *in vivo*)<sup>56</sup>. Likewise, screens for factors that alter self-renewal in ES cells and in neural stem cells can reveal chemicals or gene products that are useful for probing the biology of these cells, as well as potential therapeutics<sup>57</sup>.



**Figure 4 | Hox genes can increase the self-renewal capacity of haematopoietic cells.** Micrographs of stained blood cells are shown. Overexpression of *Hoxb4* or *Hoxa9* in mouse HSCs results in a marked increase in the number of myeloblasts (indicated by arrows, left panels) and erythroblasts after 2 weeks *in vitro*, indicating that self-renewal has increased. When the cultured cells are introduced (adoptively transferred) into irradiated animals, the cells differentiate normally. For example, neutrophils (indicated by arrows, right panels) differentiate from HSCs *in vivo*. However, *Hoxa9* overexpression leads to leukaemia, whereas *Hoxb4* overexpression does not. For the adoptive-transfer experiments, *Hoxb4*-overexpressing cells are shown at 14 weeks after transplantation; *Hoxa9*-overexpressing cells, at 3 weeks. (The large cell in the panel showing the control adoptive transfer is a megakaryocyte.) (Figure courtesy of G. Sauvageau, S. Cellot and R. Bisailon, Institute for Research in Immunology and Cancer, University of Montreal, Canada.)

Such studies have made surprising findings, such as the increase in neural stem-cell self-renewal induced by neurotransmitters.

The roles of certain small molecules in self-renewal have been studied extensively. Retinoic acid, for example, leads to alterations in *HOX* gene expression during embryogenesis and is a modifier of the WNT-mediated signalling pathway<sup>58,59</sup>. After binding to its receptor at the cell surface, retinoic acid is internalized and translocates to the nucleus, where it binds to specific nuclear hormone receptors and modulates gene transcription. A conditional knockout of the gene encoding the retinoic acid receptor RAR- $\gamma$  in mice was found to cause a prominent defect in HSC self-renewal<sup>60</sup>. Retinoic acid has also been shown to maintain HSCs in culture and can augment self-renewal in serial transplantation experiments. In addition, the development of some neuronal progenitor-cell populations is directed by timed exposure to WNT proteins and retinoic acid<sup>59</sup>. Similarly, the developmental regulation of HSC self-renewal probably involves a step at which proteins and chemical modulators interact, and it is possible that these extracellular factors affect transcription directly, by traversing the plasma membrane and translocating the nucleus, similarly to retinoic acid.

Prostaglandin E<sub>2</sub> (PGE<sub>2</sub>), a small lipid mediator, has also recently been shown to regulate HSC self-renewal during embryogenesis and can enhance HSC engraftment, as measured by competitive repopulation studies in mice<sup>56</sup>. PGE<sub>2</sub> mediates inflammatory responses and various tissue responses such as blood pressure and stomach acid production, by binding to G-protein-coupled receptors at the surface of cells and triggering the production of cyclic AMP. The concentrations of PGE<sub>2</sub> in tissues are increased during injury of tissues, and it is logical that in organs



that are attempting to repair damage, self-renewal would be coupled to factors, such as PGE<sub>2</sub>, that regulate tissue response to injury.

### A competitive transcription-factor model of self-renewal

The central role of transcription factors in stem-cell biology suggests that alterations in the levels of these factors could have marked effects on cell fate. The current view is that the switch between self-renewal and differentiation is regulated by competition between transcription-factor complexes<sup>61</sup>. Such competition may occur at the level of chromatin complexes or at the level of the transcription factors that bind to chromatin-associated factors. Cross-antagonism of transcription factors by direct binding to each other, and subsequent competition for specific target genes, could lead to a rapid shift that results in differentiation being blocked and self-renewal being maintained. For example, the transcription factor GATA1 drives the erythroid lineage, and the transcription factor PU.1 (also known as SPI1) drives the myeloid lineage<sup>61–63</sup>. These two transcription factors bind to each other, inhibiting each other's ability to interact productively with binding sites in lineage-specific genes. The relative concentration of the two factors dictates the lineage decision. If GATA1 is deficient, PU.1 will activate the myeloid programme. But, if PU.1 is deficient, GATA1 will activate the erythroid programme. A small imbalance in a transcription factor complex can therefore determine cell fate.

#### Box 1 | Conserved aspects of HSC self-renewal

The biological properties and activities that are involved in the self-renewal of HSCs are detailed below.

- Self-renewal occurs autonomously in HSCs and is also affected by the environment (or stem-cell niche).
- Self-renewal is activated by diverse signals, such as developmental regulators or certain oncogenes. The diversity of signals might make this process robust, in that a single signalling pathway is not the sole mediator. For example, Myc, NOTCH and leukaemic fusion proteins stimulate self-renewal<sup>53,96</sup>, so signalling through multiple pathways is likely to trigger a set of cellular events associated with self-renewal.
- Some of the pathways used during embryogenesis are used during adult HSC self-renewal. These include signalling pathways mediated by Delta-like, WNT proteins, BMPs, fibroblast growth factors, retinoic acid and PGE<sub>2</sub>. Some factors are required for the production of HSCs during embryogenesis but are not required for the self-renewal of HSCs at later developmental stages: for example, the transcription factors SCL and RUNX1 (ref. 93).
- HSCs do not seem to require the developmental regulators for adult stem-cell homeostasis. Conditional gene knockouts in mice have shown that developmental factors such as WNT proteins and NOTCH are not usually involved in HSC maintenance but are involved in stress-induced situations or during regeneration<sup>97,98</sup>.
- Most of the developmental pathways involved in self-renewal are conserved in many tissues and in many organisms.
- Self-renewal is linked to the cell cycle. The chromatin-associated factor BMI1 has been implicated in the self-renewal programme, and it directly regulates the transcription of the cell-cycle regulator INK4A<sup>99,100</sup>.
- Self-renewal is activated quickly. An example of this is that the transformation event in cancer cells spontaneously and rapidly activates self-renewal.
- Self-renewal can be augmented in stem cells that are already capable of self-renewal. HSCs can execute the self-renewal programme, but the addition of WNT3A<sup>24</sup>, angiopoietin-like factors<sup>75</sup> or PGE<sub>2</sub> (ref. 56) to the cells can increase the number of cells that engraft after transplantation. Most of the signalling pathways that are involved in self-renewal are not maximally stimulated *in vivo*, so it is possible to boost self-renewal.
- Self-renewal is modulated by epigenetic modification. Gene transcription is altered by modifying histones and transcription factors by processes such as phosphorylation, acetylation, ubiquitylation, SUMOylation and methylation.
- Self-renewal is linked to HOX gene expression.

The inactivation of a variety of transcription factors has been shown to inhibit HSC self-renewal, as determined by competitive transplantation assays. These factors include GATA2 (ref. 64), GFI1 (ref. 65), Myc<sup>66</sup> and SMAD4 (ref. 31). Other transcription factors that are required for the production of stem cells during embryogenesis, such as RUNX1 and SCL, are not required for self-renewal. This suggests that stem cells can detect the absence of a crucial subset of cell-specific transcription factors involved in self-renewal. Epigenetic modulation of the genes encoding these transcription factors by interacting chromatin-associated factors, and the overall state of histone modifications at these loci, is likely to be important for self-renewal.

It is well known that four transcription factors (OCT4, SOX2, Myc and KLF4) can reprogramme a fibroblast to become an induced pluripotent stem (iPS) cell, which resembles an ES cell<sup>67</sup> (see page 322). Because the original fibroblast from which the iPS cell was derived could not self-renew, this finding suggests that overexpression of crucial transcription factors can induce both pluripotency (a state in which the cell can form all the cell types of the body) and self-renewal. Transcription-factor competition could be one mechanism by which this occurs.

A competitive transcription-factor model for self-renewal would also explain the diverse events that can affect self-renewal. Self-renewal could be activated in different tissues by expression of activator transcription factors and oncogenes, coupled with depletion of repressor transcription factors. Differentiation could occur when such factors do not favour self-renewal. Mutations in these crucial regulatory genes, which encode the transcription factors, could lead to subtle changes in the stoichiometry of a competitive transcription-factor complex, favouring either self-renewal or differentiation. Alternatively, parallel developmental signalling pathways could be activated, coupling to the intrinsic cell-specific transcriptional network and thereby changing the stoichiometry of the complexes involved in self-renewal.

An attractive extension to this model is for competition to occur between different transcription factors with independent biological functions, such as competition between a transcriptional regulator of the cell cycle and a transcriptional regulator of differentiation. For example, the transcriptional inhibitor ID1 blocks differentiation along the myeloid lineage by inhibiting E-box transcription factors from binding to DNA, but the transcription factor GFI1 protects against congenital neutropenia (lack of neutrophils). Both factors have been associated with self-renewal, as determined by carrying out HSC transplantation assays in mice in which the corresponding genes have been deleted only from HSCs<sup>65,68,69</sup>. Furthermore, transcription-factor complexes that are crucial for self-renewal may integrate developmental pathways: for example, BMP-mediated signalling is known to be involved in the regulation of ID1 function<sup>69</sup>.

Although self-renewal might be regulated differently in each tissue, there are certain transcription factors, such as ZFX, that can regulate the self-renewal process in multiple tissues<sup>70</sup>. ZFX has been found to function in ES cells and HSCs to maintain self-renewal. In ES cells, this involves an anti-apoptotic pathway. After conditional inactivation of *Zfx* in HSCs, the HSCs fail to compete with wild-type bone marrow in competitive transplantation experiments. There might also be other factors that stimulate common pathways in a variety of tissues.

### Conclusion

Some of the biological aspects of the self-renewal of HSCs have now been detailed (Box 1). But researchers have long searched for the molecular features that endow stem cells with the property of self-renewal. Several years ago, two research groups put forward the idea that all populations of stem cells might have common gene-expression signatures<sup>71,72</sup>. The word 'stemness' was attached to that signature. Initially, the concept of stemness involved a set of genes that were common to neural stem cells, HSCs and ES cells. As the field grew, other stem-cell populations were studied. The small overlap of the genes in 'stemness' signatures described by many research groups suggested that the concept was invalid.

The gene-expression patterns of stem cells are likely to be distinct in all tissues, but it will be useful to compare the pathways that are necessary



and sufficient for self-renewal in a variety of organs. Studying stem cells in various species could also be useful. One possibility is that the signalling pathways that control self-renewal are conserved, with the combinatorial competition between transcription factors providing the functional basis of 'stemness.'

1. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631–644 (2008).
2. Yilmaz, O. H., Kiel, M. J. & Morrison, S. J. SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. *Blood* **107**, 924–930 (2006).
3. Merkle, F. T. & Alvarez-Buylla, A. Neural stem cells in mammalian development. *Curr. Opin. Cell Biol.* **18**, 704–709 (2006).
4. Fuller, M. T. & Spradling, A. C. Male and female *Drosophila* germline stem cells: two versions of immortality. *Science* **316**, 402–404 (2007).
5. Ailles, L. E. & Weissman, I. L. Cancer stem cells in solid tumors. *Curr. Opin. Biotechnol.* **18**, 460–466 (2007).
6. Brummendorf, T. H., Dragowska, W., Zijlman, J., Thornbury, G. & Lansdorf, P. M. Asymmetric cell divisions sustain long-term hematopoiesis from single-sorted human fetal liver cells. *J. Exp. Med.* **188**, 1117–1124 (1998).
7. Takano, H., Ema, H., Sudo, K. & Nakauchi, H. Asymmetric division and lineage commitment at the level of hematopoietic stem cells: inference from differentiation in daughter cell and granddaughter cell pairs. *J. Exp. Med.* **199**, 295–302 (2004).
8. Beckmann, J., Scheitza, S., Wernet, P., Fischer, J. C. & Giebel, B. Asymmetric cell division within the human hematopoietic stem and progenitor cell compartment: identification of asymmetrically segregating proteins. *Blood* **109**, 5494–5501 (2007).
9. Yamashita, Y. M., Mahowald, A. P., Perl, J. R. & Fuller, M. T. Asymmetric inheritance of mother versus daughter centrosome in stem cell division. *Science* **315**, 518–521 (2007).
10. Brinster, R. L. Male germline stem cells: from mice to men. *Science* **316**, 404–405 (2007).
11. Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635–648 (2004).
12. Stingl, J. et al. Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
13. Radtke, F. & Clevers, H. Self-renewal and cancer of the gut: two sides of a coin. *Science* **307**, 1904–1909 (2005).
14. Collins, C. A. & Partridge, T. A. Self-renewal of the adult skeletal muscle satellite cell. *Cell Cycle* **4**, 1338–1341 (2005).
15. Dor, Y., Brown, J., Martinez, O. I. & Melton, D. A. Adult pancreatic  $\beta$ -cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41–46 (2004).
16. Pan, G. & Thomson, J. A. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res.* **17**, 42–49 (2007).
17. Nakagawa, T., Nabeshima, Y. & Yoshida, S. Functional identification of the actual and potential stem cell compartments in mouse spermatogenesis. *Dev. Cell* **12**, 195–206 (2007).
- This paper shows that transient amplifying cells in the testes can acquire stem-cell activity and self-renewal.**
18. Fuchs, E. Skin stem cells: rising to the surface. *J. Cell Biol.* **180**, 273–284 (2008).
19. Bowie, M. B. et al. Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. *Proc. Natl Acad. Sci. USA* **104**, 5878–5882 (2007).
- The paper found that the self-renewal capacity of fetal HSCs differs from that of adult bone-marrow-derived stem cells.**
20. Kim, I., Saunders, T. L. & Morrison, S. J. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. *Cell* **130**, 470–483 (2007).
21. Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nature Immunol.* **9**, 129–136 (2008).
22. Ober, E. A., Verkade, H., Field, H. A. & Stainier, D. Y. Mesodermal Wnt2b signalling positively regulates liver specification. *Nature* **442**, 688–691 (2006).
23. McMahon, A. P. & Bradley, A. The *Wnt-1 (int-1)* proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**, 1073–1085 (1990).
24. Reya, T. et al. A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–414 (2003).
25. Lai, K., Kaspar, B. K., Gage, F. H. & Schaffer, D. V. Sonic hedgehog regulates adult neural progenitor proliferation *in vitro* and *in vivo*. *Nature Neurosci.* **6**, 21–27 (2003).
26. Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. & Fuller, M. T. Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294**, 2542–2545 (2001).
27. Molofsky, A. V. et al. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
28. Koch, U. et al. Simultaneous loss of  $\beta$ - and  $\gamma$ -catenin does not perturb hematopoiesis or lymphopoiesis. *Blood* **111**, 160–164 (2008).
- This study shows that the WNT-mediated signalling pathway is not required for HSC homeostasis in adulthood, even though WNT proteins are sufficient for the proliferation of HSCs.**
29. Zhao, C. et al. Loss of  $\beta$ -catenin impairs the renewal of normal and CML stem cells *in vivo*. *Cancer Cell* **12**, 528–541 (2007).
30. Stoick-Cooper, C. I. et al. Distinct Wnt signaling pathways have opposing roles in appendage regeneration. *Development* **134**, 479–489 (2007).
31. Karlsson, G. et al. Smad4 is critical for self-renewal of hematopoietic stem cells. *J. Exp. Med.* **204**, 467–474 (2007).
32. Whitehead, G. G., Makino, S., Lien, C. L. & Keating, M. T. *fgr20* is essential for initiating zebrafish fin regeneration. *Science* **310**, 1957–1960 (2005).
- This study shows that *Fgf20* is a dedicated growth factor for tail-fin regeneration in zebrafish, suggesting that each tissue might have its own set of growth conditions for regeneration.**

33. Bowie, M. B., Kent, D. G., Copley, M. R. & Eaves, C. J. Steel factor responsiveness regulates the high self-renewal phenotype of fetal hematopoietic stem cells. *Blood* **109**, 5043–5048 (2007).
34. Kirstetter, P., Anderson, K., Porse, B. T., Jacobsen, S. E. & Nerlov, C. Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. *Nature Immunol.* **7**, 1048–1056 (2006).
35. Scheller, M. et al. Hematopoietic stem cell and multilineage defects generated by constitutive  $\beta$ -catenin activation. *Nature Immunol.* **7**, 1037–1047 (2006).
36. Bondos, S. Variations on a theme: Hox and Wnt combinatorial regulation during animal development. *Sci. STKE* **2006**, pe38 (2006).
- References 35 and 36 show that constitutive activation of  $\beta$ -catenin leads to a block in differentiation and indicates that repeated activation of pathways that increase HSC number can lead to stem-cell exhaustion.**
37. Pilon, N. et al. *Cdx4* is a direct target of the canonical Wnt pathway. *Dev. Biol.* **289**, 55–63 (2006).
38. Davidson, A. J. et al. *cdx4* mutants fail to specify blood progenitors and can be rescued by multiple *hox* genes. *Nature* **425**, 300–306 (2003).
39. Sauvageau, G. et al. Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations *in vitro* and *in vivo*. *Genes Dev.* **9**, 1753–1765 (1995).
40. Thorsteinsdottir, U. et al. Overexpression of the myeloid leukemia-associated *Hoxa9* gene in bone marrow cells induces stem cell expansion. *Blood* **99**, 121–129 (2002).
41. Magnusson, M. et al. HOXA10 is a critical regulator for hematopoietic stem cells and erythroid/megakaryocyte development. *Blood* **109**, 3687–3696 (2007).
42. Schnabel, C. A., Jacobs, Y. & Cleary, M. L. HoxA9-mediated immortalization of myeloid progenitors requires functional interactions with TALE cofactors Pbx and Meis. *Oncogene* **19**, 608–616 (2000).
43. Magnusson, M., Brun, A. C., Lawrence, H. J. & Karlsson, S. Hoxa9/hoxb3/hoxb4 compound null mice display severe hematopoietic defects. *Exp. Hematol.* **35**, 1421–1428 (2007).
44. Park, I. K. et al. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302–305 (2003).
45. Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255–260 (2003).
46. Lessard, J. et al. Functional antagonism of the Polycomb-group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev.* **13**, 2691–2703 (1999).
47. Oguro, H. et al. Differential impact of Ink4a and Arf on hematopoietic stem cells and their bone marrow microenvironment in Bmi1-deficient mice. *J. Exp. Med.* **203**, 2247–2253 (2006).
48. Bruggeman, S. W. et al. Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. *Genes Dev.* **19**, 1438–1443 (2005).
- References 47 and 48 show that the chromatin-associated factor Bmi1 controls the cell cycle of stem cells by altering expression of INK4A and ARF.**
49. Datta, S. et al. Bmi-1 cooperates with H-Ras to transform human mammary epithelial cells via dysregulation of multiple growth-regulatory pathways. *Cancer Res.* **67**, 10286–10295 (2007).
50. Liu, S. et al. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res.* **66**, 6063–6071 (2006).
51. Molofsky, A. V., He, S., Bydon, M., Morrison, S. J. & Pardoll, R. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16<sup>INK4a</sup> and p19<sup>Arf</sup> senescence pathways. *Genes Dev.* **19**, 1432–1437 (2005).
52. Tateishi, K. et al. Dysregulated expression of stem cell factor Bmi1 in precancerous lesions of the gastrointestinal tract. *Clin. Cancer Res.* **12**, 6960–6966 (2006).
53. Krivtsov, A. V. et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* **442**, 818–822 (2006).
54. Ernst, P. et al. Definitive hematopoiesis requires the mixed-lineage leukemia gene. *Dev. Cell* **6**, 437–443 (2004).
55. Tadokoro, Y., Ema, H., Okano, M., Li, E. & Nakauchi, H. *De novo* DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J. Exp. Med.* **204**, 715–722 (2007).
56. North, T. E. et al. Prostaglandin E<sub>2</sub> regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011 (2007).
57. Diamandis, P. et al. Chemical genetics reveals a complex functional ground state of neural stem cells. *Nature Chem. Biol.* **3**, 268–273 (2007).
58. Shiotsugu, J. et al. Multiple points of interaction between retinoic acid and FGF signaling during embryonic axis formation. *Development* **131**, 2653–2667 (2004).
59. Nordstrom, U., Maier, E., Jessell, T. M. & Edlund, T. An early role for WNT signaling in specifying neural patterns of *Cdx* and *Hox* gene expression and motor neuron subtype identity. *PLoS Biol.* **4**, e252 (2006).
60. Purton, L. E. et al. RAR $\gamma$  is critical for maintaining a balance between hematopoietic stem cell self-renewal and differentiation. *J. Exp. Med.* **203**, 1283–1293 (2006).
61. Nerlov, C. & Graf, T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.* **12**, 2403–2412 (1998).
- This paper reports that the overexpression of transcription factors can lead to an altered cell-fate programme.**
62. Reikhtman, N., Radparvar, F., Evans, T. & Skoultschi, A. I. Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes Dev.* **13**, 1398–1411 (1999).
- This study found that competition between transcription factors can alter cell-fate decisions.**
63. Galloway, J. L., Wingert, R. A., Thisse, C., Thisse, B. & Zon, L. I. Loss of *gata1* but not *gata2* converts erythropoiesis to myelopoiesis in zebrafish embryos. *Dev. Cell* **8**, 109–116 (2005).
64. Rodriguez, N. P. et al. Haploinsufficiency of GATA-2 perturbs adult hematopoietic stem-cell homeostasis. *Blood* **106**, 477–484 (2005).
65. Zeng, H., Yucel, R., Kosan, C., Klein-Hitpass, L. & Moroy, T. Transcription factor Gfi1 regulates self-renewal and engraftment of hematopoietic stem cells. *EMBO J.* **23**, 4116–4125 (2004).



66. Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
67. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
68. Jankovic, V. *et al.* Id1 restrains myeloid commitment, maintaining the self-renewal capacity of hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 1260–1265 (2007).
69. Hollnagel, A., Oehlmann, V., Heymer, J., Ruther, U. & Nordheim, A. *Id* genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J. Biol. Chem.* **274**, 19838–19845 (1999).
70. Galan-Cardad, J. M. *et al.* Zfx controls the self-renewal of embryonic and hematopoietic stem cells. *Cell* **129**, 345–357 (2007).  
This paper shows that the transcription factor ZFX is required for the self-renewal of several stem-cell populations.
71. Ramalho-Santos, M., Yoon, S., Matsuzaki, Y., Mulligan, R. C. & Melton, D. A. 'Stemness': transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597–600 (2002).
72. Ivanova, N. B. *et al.* A stem cell molecular signature. *Science* **298**, 601–604 (2002).
73. Varnum-Finney, B. *et al.* The Notch ligand, Jagged-1, influences the development of primitive hematopoietic precursor cells. *Blood* **91**, 4084–4091 (1998).
74. Stier, S., Cheng, T., Dombkowski, D., Carlesso, N. & Scadden, D. T. Notch1 activation increases hematopoietic stem cell self-renewal *in vivo* and favors lymphoid over myeloid lineage outcome. *Blood* **99**, 2369–2378 (2002).
75. Zhang, C. C. *et al.* Angiopoietin-like proteins stimulate *ex vivo* expansion of hematopoietic stem cells. *Nature Med.* **12**, 240–245 (2006).
76. Petit-Cocault, L., Valle-Challier, C., Fleury, M., Peault, B. & Souyri, M. Dual role of Mpl receptor during the establishment of definitive hematopoiesis. *Development* **134**, 3031–3040 (2007).
77. Janzen, V. *et al.* Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16<sup>INK4a</sup>. *Nature* **443**, 421–426 (2006).
78. Yuan, Y., Shen, H., Franklin, D. S., Scadden, D. T. & Cheng, T. *In vivo* self-renewing divisions of hematopoietic stem cells are increased in the absence of the early G1-phase inhibitor, p18<sup>INK4C</sup>. *Nature Cell Biol.* **6**, 436–442 (2004).
79. Walkley, C. R., Fero, M. L., Chien, W. M., Purton, L. E. & McArthur, G. A. Negative cell-cycle regulators cooperatively control self-renewal and differentiation of haematopoietic stem cells. *Nature Cell Biol.* **7**, 172–178 (2005).
80. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21<sup>Cip1/Waf1</sup>. *Science* **287**, 1804–1808 (2000).
81. Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475–482 (2006).
82. Zhang, J. *et al.* PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441**, 518–522 (2006).
83. TeKippe, M., Harrison, D. E. & Chen, J. Expansion of hematopoietic stem cell phenotype and activity in Trp53-null mice. *Exp. Hematol.* **31**, 521–527 (2003).
84. Nakamura, T., Largaespada, D. A., Shaughnessy, J. D., Jenkins, N. A. & Copeland, N. G. Cooperative activation of *Hoxa* and *Pbx1*-related genes in murine myeloid leukaemias. *Nature Genet.* **12**, 149–153 (1996).
85. Frohling, S., Scholl, C., Bansal, D. & Huntly, B. J. *HOX* gene regulation in acute myeloid leukemia: CDX marks the spot? *Cell Cycle* **6**, 2241–2245 (2007).
86. Wong, P., Iwasaki, M., Somervaille, T. C., So, C. W. & Cleary, M. L. Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. *Genes Dev.* **21**, 2762–2774 (2007).
87. Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* **18**, 2336–2341 (2004).
88. Passegue, E., Wagner, E. F. & Weissman, I. L. JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells. *Cell* **119**, 431–443 (2004).
89. Iwasaki, H. *et al.* Distinctive and indispensable roles of PU.1 in maintenance of hematopoietic stem cells and their differentiation. *Blood* **106**, 1590–1600 (2005).
90. Sandberg, M. L. *et al.* c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. *Dev. Cell* **8**, 153–166 (2005).
91. Rebel, V. I. *et al.* Distinct roles for CREB-binding protein and p300 in hematopoietic stem cell self-renewal. *Proc. Natl Acad. Sci. USA* **99**, 14789–14794 (2002).
92. Gowney, J. D. *et al.* Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* **106**, 494–504 (2005).
93. Mikkola, H. K. *et al.* Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature* **421**, 547–551 (2003).
94. Kamminga, L. M. *et al.* The Polycomb group gene *Ezh2* prevents hematopoietic stem cell exhaustion. *Blood* **107**, 2170–2179 (2006).
95. Ohta, H. *et al.* Polycomb group gene *rae28* is required for sustaining activity of hematopoietic stem cells. *J. Exp. Med.* **195**, 759–770 (2002).
96. Varnum-Finney, B. *et al.* Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. *Nature Med.* **6**, 1278–1281 (2000).
97. Mancini, S. J. *et al.* Jagged1-dependent Notch signaling is dispensable for hematopoietic stem cell self-renewal and differentiation. *Blood* **105**, 2340–2342 (2005).
98. Congdon, K. L. *et al.* Activation of Wnt signaling in hematopoietic regeneration. *Stem Cells* doi:10.1634/stemcells.2007-0768 (in the press).
99. Kotake, Y. *et al.* pRB family proteins are required for H3K27 trimethylation and Polycomb repression complexes binding to and silencing p16<sup>INK4a</sup> tumor suppressor gene. *Genes Dev.* **21**, 49–54 (2007).
100. Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the *INK4A-ARF* locus and are disassociated in senescent cells. *Genes Dev.* **21**, 525–530 (2007).

**Acknowledgements** I thank G. Sauvageau, S. Cellot and R. Bisailon, R. Moon and C. Nerlov for providing figures. Work in my laboratory was supported by the Howard Hughes Medical Institute and grants from the National Institutes of Health.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The author declares competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to the author ([zon@enders.tch.harvard.edu](mailto:zon@enders.tch.harvard.edu)).



# Wound repair and regeneration

Geoffrey C. Gurtner<sup>1</sup>, Sabine Werner<sup>2</sup>, Yann Barrandon<sup>3,4</sup> & Michael T. Longaker<sup>1</sup>

**The repair of wounds is one of the most complex biological processes that occur during human life. After an injury, multiple biological pathways immediately become activated and are synchronized to respond. In human adults, the wound repair process commonly leads to a non-functioning mass of fibrotic tissue known as a scar. By contrast, early in gestation, injured fetal tissues can be completely recreated, without fibrosis, in a process resembling regeneration. Some organisms, however, retain the ability to regenerate tissue throughout adult life. Knowledge gained from studying such organisms might help to unlock latent regenerative pathways in humans, which would change medical practice as much as the introduction of antibiotics did in the twentieth century.**

In the moments after an injury occurs, various intracellular and intercellular pathways must be activated and coordinated if tissue integrity and homeostasis are to be restored. Cellular components of the immune system, the blood coagulation cascade and the inflammatory pathways are activated in addition. Many types of cell — including immune cells (neutrophils, monocytes, lymphocytes and dendritic cells), endothelial cells, keratinocytes and fibroblasts — undergo marked changes in gene expression and phenotype, leading to cell proliferation, differentiation and migration<sup>1,2</sup>. If this response is successful and the injury does not result in the demise of the organism, these processes must be shut down in a precise sequence in the ensuing days as recovery progresses.

Given the complexity of the wound repair process, it is remarkable that it rarely becomes uncontrolled and that malignant transformation is an uncommon event in the wound environment<sup>3,4</sup>. For most injuries, repair results in once functional tissue becoming a patch of cells (mainly fibroblasts) and disorganized extracellular matrix (mainly collagen) that is commonly referred to as a scar. Surprisingly, in some eukaryotic organisms, the response to injury can completely recapitulate the original tissue architecture, through the poorly understood process of regeneration. Humans have this ability during prenatal development, but it is lost during adult life<sup>5</sup>. How regeneration occurs and why humans lose this ability remain a mystery<sup>6</sup>. In this review, we provide an overview of how multicellular organisms respond to tissue loss, highlighting areas in which developmental pathways have been used to foster tissue regeneration. Recent advances in epithelial stem-cell biology and developmental biology have begun to elucidate the pathways that need to be reactivated to effect regeneration in humans.

## Clinical burden of fibrosis

In general, the wound repair process occurs in almost all tissues after exposure to almost any destructive stimulus. Thus, the sequence of events that follows a myocardial infarction (heart attack), for example, is remarkably similar to that following a spinal-cord injury, a burn or a gunshot wound, despite the different types of insult and the different organs affected. Likewise, scar formation that occurs during wound repair leads to similar tissue dysfunction wherever it takes place.

In the case of myocardial infarction, the formation of myocardial scar tissue is thought to result in congestive heart failure (a condition in which

the heart cannot supply the body's tissues with enough blood) and/or abnormal heart rhythms (arrhythmias) (see page 322), which together account for nearly 100,000 deaths each year in the United States alone<sup>7</sup>. In addition, cirrhosis of the liver and some forms of fibrosis of the lungs are thought to result from fibrotic responses to toxin-mediated injury<sup>8,9</sup>. Interestingly, in other circumstances, the liver is one of the few organs in the human body that can regenerate up to 70% of itself without scar formation. Why the liver's regenerative capacity manifests only in some cases remains incompletely understood. A leading hypothesis is that the immune system is involved in the switch between regeneration and fibrotic healing, because human fetuses, which heal without scarring, have immature immune systems<sup>10</sup>.

Healing by fibrosis, instead of regeneration, places a huge burden on public health. The total economic cost of diseases that result from fibrosis is difficult to calculate precisely but is of the order of tens of billions of dollars<sup>11</sup>. Importantly, dysfunctional healing often causes lifelong disability, which has a significant economic impact<sup>2</sup>. Thus, if fibrotic healing processes can be transformed into regenerative ones, in which the original tissues are restored, this would considerably improve human health.

## Wound repair across phylogeny

The ability to respond to injury and to repair tissue is a fundamental property of all multicellular organisms. But there is tremendous diversity in how this process occurs. Studying wound repair in various phyla could improve our understanding of wound repair in humans and might help to identify molecules or pathways that can be targeted to restore the lost regenerative capacity.

The most primitive multicellular organism and the common ancestor of modern multicellular organisms is the sponge. Although wound repair has not been studied extensively in sponges, recent work on cellular patterning in these organisms has provided insight into the origin of embryonic patterning in metazoans. Patterning refers to the three-dimensional orientation of cells in an organism and is probably required for tissue regeneration. It is generally not thought to be involved in tissue repair with scarring. Sponges are simple organisms without a body axis, multiple layers of cells, or tissues<sup>12–14</sup>. Adult sponges have highly adaptable body shapes and lack an anteroposterior axis of symmetry,

<sup>1</sup>Division of Plastic and Reconstructive Surgery, Department of Surgery, Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, 257 Campus Drive, Stanford, California 94305-5148, USA. <sup>2</sup>Institute of Cell Biology, Department of Biology, Swiss Federal Institute of Technology (ETH), Schafmattstrasse 18, HPM D42, CH-8093 Zürich, Switzerland. <sup>3</sup>Centre Hospitalier Universitaire Vaudois, Chirurgie Expérimentale, Pavillon 4, CH-1011 Lausanne, Switzerland. <sup>4</sup>École Polytechnique Fédérale Lausanne, School of Life Sciences/LDCS, Station 15, CH-1015 Lausanne, Switzerland.



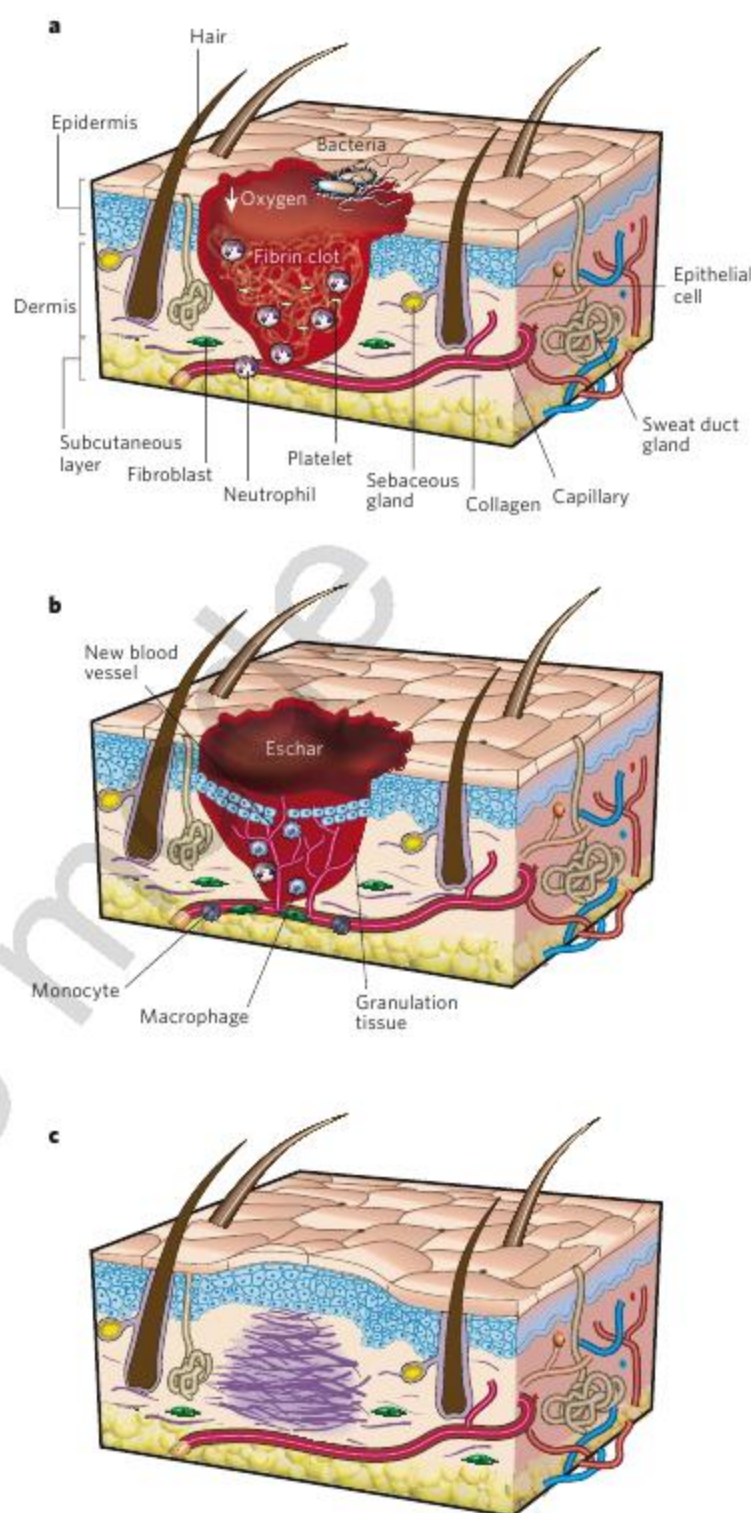
but embryonic and larval sponges have both an anteroposterior axis and radial symmetry. Certain soluble factors — WNT proteins and transforming growth factor- $\beta$  (TGF- $\beta$ ) — together are required to pattern the embryonic sponge cells, and a hedgehog-like cell-surface signal, *hedgling*<sup>15</sup>, is also produced in overlapping patterns with the WNT proteins and TGF- $\beta$ .

Similarly, in *Drosophila melanogaster*, morphological events that occur during development, such as dorsal closure<sup>16</sup> and tracheal fusion<sup>17</sup>, have a remarkable visual similarity to wound repair in humans. Unlike sponges, *D. melanogaster* has a complex, craniocaudal asymmetry in the adult stage. In addition, in *D. melanogaster* embryos, the same molecular machinery is used in these morphological developmental events and to close an epithelial gap in a wound<sup>18</sup>. That these intracellular signalling pathways and cell-adhesion pathways are highly conserved across species and components are present after injury in other multicellular organisms, including humans, indicates that the patterning cues for multicellular organization are present in the wound environment in adults but are not fully functional<sup>19</sup>.

To examine the involvement of developmental pathways in wound repair more directly, morphogenetic events involving tissue movement have been used as models of wound repair. The two most commonly used models are dorsal closure in *D. melanogaster* embryos and ventral enclosure in *Caenorhabditis elegans*, both of which involve the elimination of epithelial gaps. In *D. melanogaster*, it seems that JUN amino-terminal kinase (JNK) signalling and the subsequent activation of the transcription factor activator protein 1 (AP1) in leading-edge epithelial cells, together with tight regulation of the associated cell division and cell-cell adhesion, are crucial for successful closure<sup>20</sup>. As discussed later, this process is analogous to the re-epithelialization (the recreation of an intact keratinocyte layer) of mammalian wounds, again emphasizing the common foundation of tissue repair across phylogeny.

Wound repair in amphibians has been the focus of recent excitement because of the relatively close relationship between these organisms and mammals and because of their remarkable ability to regenerate amputated appendages through the formation of a blastema, which is a mass of undifferentiated cells capable of regeneration. After wounding, differentiated cells in the mature tissues that surround the amputated region dedifferentiate into mononuclear blastemal cells that can proliferate and have the potential to differentiate into multiple cell types<sup>21,22</sup>. These cells then regenerate the limb through a process that mirrors embryonic development. Nerve stimulation is important for maintaining this regenerative state<sup>23</sup>. Recently, it has been shown in salamanders that nAG, a member of the anterior-gradient protein family, is produced by the Schwann cells of transected nerves and interacts with Prod 1 on the surface of blastemal cells, promoting their proliferation<sup>24</sup>. However, if the nerves proximal to the blastema (that is, near the spinal cord) are transected, then the blastema fails to undergo limb regeneration, and the cells fail to redifferentiate; therefore, wound repair occurs instead of limb regeneration. Similarly, planarians are simple animals that lack complex organ systems but can fully regenerate an intact organism from a minor remnant (as little as 1/289) of the original organism. Once again, undifferentiated cells (neoblasts) collect at the injury site and form a blastema, which can regenerate organ systems and tissues<sup>25</sup>. In planarians, it is thought that up to 20% of all cells are committed to the regeneration of adult tissues.

Mammals have retained much of the molecular machinery used by organisms such as salamanders, but their regenerative potential is only limited. In part, this seems to result from the rapid interposition of fibrotic tissue, which prevents subsequent tissue regeneration. For example, when the spinal cord of mice is injured, neurons begin to grow, but glial cells at the site of injury stimulate scar formation, preventing recovery. However, if the mouse spinal cord is cut so that scar formation is hindered, the neurons reconnect<sup>26,27</sup>. This rapid interposition of scar tissue probably confers a survival advantage by preventing infectious microorganisms from invading the wound and by inhibiting the continued mechanical deformation of larger tissues (a process that could compound the initial insult). To manipulate the wound repair process



**Figure 1 | Classic stages of wound repair.** There are three classic stages of wound repair: inflammation (a), new tissue formation (b) and remodelling (c). **a, Inflammation.** This stage lasts until about 48 h after injury. Depicted is a skin wound at about 24–48 h after injury. The wound is characterized by a hypoxic (ischaemic) environment in which a fibrin clot has formed. Bacteria, neutrophils and platelets are abundant in the wound. Normal skin appendages (such as hair follicles and sweat duct glands) are still present in the skin outside the wound. **b, New tissue formation.** This stage occurs about 2–10 days after injury. Depicted is a skin wound at about 5–10 days after injury. An eschar (scab) has formed on the surface of the wound. Most cells from the previous stage of repair have migrated from the wound, and new blood vessels now populate the area. The migration of epithelial cells can be observed under the eschar. **c, Remodelling.** This stage lasts for a year or longer. Depicted is a skin wound about 1–12 months after repair. Disorganized collagen has been laid down by fibroblasts that have migrated into the wound. The wound has contracted near its surface, and the widest portion is now the deepest. The re-epithelialized wound is slightly higher than the surrounding surface, and the healed region does not contain normal skin appendages.



in mammals so that it shifts towards regeneration will probably require the ability to slow the rapid fibrotic response so that multipotent cells such as stem or progenitor cells can allow tissue regeneration rather than scar formation.

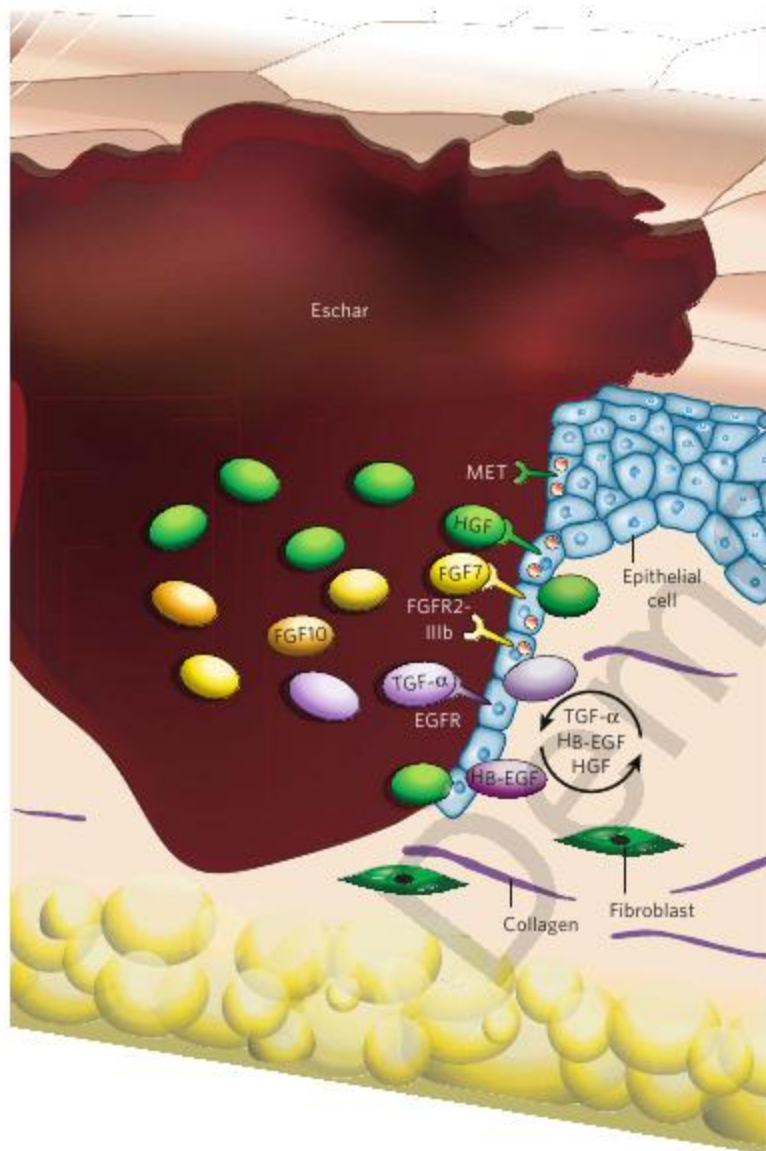
### Classic stages of wound repair

In all organ systems, the normal mammalian response to injury occurs in three overlapping but distinct stages: inflammation, new tissue formation, and remodelling (Fig. 1). The first stage of wound repair — inflammation — occurs immediately after tissue damage, and components of the coagulation cascade, inflammatory pathways and immune system are needed to prevent ongoing blood and fluid losses, to remove dead and devitalized (dying) tissues and to prevent infection. Haemostasis is achieved initially by the formation of a platelet plug, followed by a fibrin matrix, which becomes the scaffold for infiltrating cells. Neutrophils are then recruited to the wound in response to the activation of complement, the degranulation of platelets and the

products of bacterial degradation<sup>28</sup> (Fig. 1a). After 2–3 days, monocytes appear in the wound and differentiate into macrophages. Macrophages are thought to be crucial for coordinating later events in the response to injury, but the importance of neutrophils and macrophages in wound repair is incompletely understood. Recent data suggest, however, that a deficiency in either cell type can be compensated for by the redundancy in the inflammatory response<sup>29</sup>. In the absence of both cell types, the repair of small wounds can still occur, and the scarring response is even less<sup>30</sup>.

The second stage of wound repair — new tissue formation — occurs 2–10 days after injury and is characterized by cellular proliferation and migration of different cell types. The first event is the migration of keratinocytes over the injured dermis (the inner layer of the skin) (Fig. 1b). New blood vessels then form (a process known as angiogenesis), and the sprouts of capillaries associated with fibroblasts and macrophages replace the fibrin matrix with granulation tissue, which forms a new substrate for keratinocyte migration at later stages of the repair process. The keratinocytes that are behind the leading edge proliferate and mature and, finally, restore the barrier function of the epithelium. The most important positive regulators of angiogenesis<sup>31</sup> are vascular endothelial growth factor A (VEGFA) and fibroblast growth factor 2 (FGF2; also known as bFGF). For example, the application of VEGFA alone to wounds in an animal model of diabetes (in which wound repair is dysregulated) can normalize healing<sup>32</sup>. Angiogenesis can also result from the recruitment of bone-marrow-derived endothelial progenitor cells, although the magnitude of this contribution is small — at least in non-ischaemic wounds (in which the concentration of oxygen is normal)<sup>33</sup> (Fig. 1b). In the later part of this stage, fibroblasts, which are attracted from the edge of the wound or from the bone marrow are stimulated by macrophages, and some differentiate into myofibroblasts<sup>34</sup>. Myofibroblasts are contractile cells that, over time, bring the edges of a wound together. Fibroblasts and myofibroblasts interact and produce extracellular matrix, mainly in the form of collagen, which ultimately forms the bulk of the mature scar<sup>35</sup>.

The third stage of wound repair — remodelling — begins 2–3 weeks after injury and lasts for a year or more. During this stage, all of the processes activated after injury wind down and cease. Most of the endothelial cells, macrophages and myofibroblasts undergo apoptosis (that is, programmed cell death) or exit from the wound, leaving a mass that contains few cells and consists mostly of collagen and other extracellular-matrix proteins (Fig. 1c). Epithelial–mesenchymal interactions probably continuously regulate skin integrity and homeostasis<sup>36</sup>. And there must be additional feedback loops to maintain the other cell types in the skin. In addition, over 6–12 months, the acellular matrix is actively remodelled from a mainly type III collagen backbone to one predominantly composed of type I collagen<sup>37</sup>. This process is carried out by matrix metalloproteinases that are secreted by fibroblasts, macrophages and endothelial cells, and it strengthens the repaired tissue. However, the tissue never regains the properties of uninjured skin<sup>38</sup>. Interestingly, vertebrates such as zebrafish (*Danio rerio*) and *C. elegans* do not produce either of these collagen types; their extracellular matrix consists entirely of type VI and type XVIII collagens<sup>39</sup>. This finding suggests a degree of evolutionary plasticity that is not observed in the earlier stages of wound repair.



**Figure 2 | Wound re-epithelialization.** Half of an excisional wound in skin is shown at about 2 days after injury. Growth factors that are known to stimulate wound re-epithelialization are depicted: hepatocyte growth factor (HGF), which binds to MET; fibroblast growth factor 7 (FGF7) and FGF10, which bind to the IIIb isoform of FGF receptor 2 (FGFR2-IIIb), and ligands of the epidermal growth factor receptor (EGFR), such as transforming growth factor- $\alpha$  (TGF- $\alpha$ ) and heparin-binding epidermal growth factor (HB-EGF). The signalling pathways initiated by these interactions activate the transcription factors signal transducer and activator of transcription 3 (STAT3) and AP1 proteins (pink circles), which help to regulate wound re-epithelialization. This process is further promoted by peroxisome-proliferator-activated receptors (PPARs), which are most probably activated by non-saturated fatty acids. The loop indicates autocrine stimulation.

### Molecular mechanisms of wound repair

Following the observation that genes that are regulated in response to skin injury are functionally important for the wound repair process, DNA-microarray studies were carried out to identify such genes across the genome<sup>40,41</sup>. These studies showed that the gene-expression pattern of healing skin wounds strongly resembles that of highly malignant tumours<sup>42</sup>, highlighting the importance of functional genomics studies for research into wound repair and cancer. The generation and analysis of genetically modified mice provided insight into the roles of some of the genes regulated during wound repair<sup>28</sup>. In addition, the observed similarities between the cell movements that occur during dorsal closure in *D. melanogaster* embryos and the healing of mammalian skin



**Table 1 | Soluble mediators of re-epithelialization**

Ligand	Receptor	Type of receptor	Signalling proteins	Role in re-epithelialization	References
HGF	MET	Receptor tyrosine kinase	Unknown, possibly ERK1 and ERK2, AKT, GAB1, PAK1 and/or PAK2	Stimulation of keratinocyte migration and probably proliferation	43
FGF7, FGF10 and FGF22	FGFR2-IIIb, possibly FGFR1-IIIb	Receptor tyrosine kinase	Unknown, possibly ERK1, ERK2, AKT and/or STAT3	Stimulation of keratinocyte proliferation and migration	44–46
Heparin-binding EGF and other EGF-family members	EGFR (also known as ERBB1), possibly ERBB2, ERBB3 and/or ERBB4	Receptor tyrosine kinase	Unknown, possibly ERK1 and ERK2, AKT and/or STAT3	Stimulation of keratinocyte proliferation and migration	30, 47
TGF- $\beta$	TGF- $\beta$ receptor I and TGF- $\beta$ receptor II	Receptor serine/threonine kinase	SMAD3 and others, including SMAD2 and MAPK	Inhibition of keratinocyte proliferation and survival	30, 51, 52
Acetylcholine	M3 receptor	G-protein-coupled receptor	Ca <sup>2+</sup> -dependent guanylyl cyclase, cyclic GMP and PKG, leading to inhibition of RHO	Inhibition of keratinocyte migration	54
	M4 receptor	G-protein-coupled receptor	Adenylyl cyclase, cyclic AMP and PKA, leading to activation of RHO	Stimulation of keratinocyte migration	54
Catecholamines, including adrenaline	$\beta_2$ -Adrenoceptor	G-protein-coupled receptor	Activation of phosphatase PP2A, resulting in dephosphorylation and inhibition of ERK1 and ERK2	Inhibition of keratinocyte migration	55
Polyunsaturated fatty acids	PPAR- $\alpha$ and PPAR- $\beta$ *	Nuclear receptor	Direct activation of target genes by binding to the promoter/enhancer of these genes	Stimulation of keratinocyte migration and survival	56–58

EGF, epidermal growth factor; EGFR, EGF receptor; ERK, extracellular-signal-regulated kinase; FGF, fibroblast growth factor; FGFR1-IIIb, IIIb isoform of FGF receptor 1; GAB1, growth-factor-receptor-bound protein 2 (GRB2)-associated binding protein 1; HGF, hepatocyte growth factor; M3, muscarinic receptor subtype 3; PAK, p21-activated kinase; PKA, cyclic-AMP-dependent protein kinase; PKG, cyclic-GMP-dependent protein kinase; PPAR, peroxisome-proliferator-activated receptor; SMAD3, SMAD-family member 3; STAT3, signal transducer and activator of transcription 3; TGF- $\beta$ , transforming growth factor- $\beta$ . \*PPAR- $\beta$  ligands might be fatty acids.

wounds, together with the establishment of wound repair models in *D. melanogaster*<sup>20</sup>, have allowed the genetically tractable *D. melanogaster* system to be used for identifying and functionally characterizing the molecules involved in wound repair, particularly in re-epithelialization. Many of the steps in wound repair are poorly characterized at the molecular level at present, so in this section we discuss recent data that reveal the underlying molecular mechanisms of re-epithelialization, a major event in the phase of new tissue formation (second stage).

The proteins involved in re-epithelialization include various extracellular-matrix proteins and their receptors, proteases (including matrix metalloproteinases), cytoskeletal proteins, and enzymes that regulate the cellular redox balance. The functions of these proteins in wound repair have been described in a recent review<sup>43</sup>. Other proteins that are known to be involved include growth factors and hormones, and we focus on the role of these proteins and their downstream targets in the repair of the injured epidermis (the outer layer of the skin) and the dermis (Fig. 2 and Table 1).

### Peptide growth factors

One molecule that is particularly important in re-epithelialization is hepatocyte growth factor (HGF), which exerts its function by binding to and activating MET, a receptor tyrosine kinase<sup>31,44</sup>. Interestingly, mice in which the gene encoding MET had been deleted from keratinocytes showed strongly delayed re-epithelialization in response to skin wounds, and the cells that eventually covered the wounds in these mice were found to have escaped the recombination event in which *Met* was deleted and therefore still expressed MET<sup>44</sup>. This result was unexpected, because it was known that other growth factors are involved in re-epithelialization, but it is clear that these factors cannot compensate for a lack of HGF-mediated signalling. Other growth factors that positively regulate re-epithelialization include members of the FGF family and the epidermal growth factor (EGF) family. In addition, certain peptide growth factors, most notably TGF- $\beta$ , negatively regulate re-epithelialization.

In terms of the FGF family, ligands for the IIIb variant of FGF receptor 2 (FGFR2-IIIb), in particular, contribute to wound repair. This is shown by the strong delay in re-epithelialization that occurs in transgenic mice expressing a dominant-negative mutant of FGFR2-IIIb in keratinocytes. The mutant receptor binds to FGF but cannot transduce the signal<sup>45</sup>. The most important FGFR2-IIIb ligands, the functions of which were abrogated in these mice, are likely to be FGF7 and FGF10 (ref. 46). In support of this idea, mice that lack dendritic epidermal T cells, which are a potent source of FGF7 and FGF10 (ref. 47), show decreased keratinocyte proliferation and wound closure after injury.

Many of the ligands for the EGF receptor (EGFR) are expressed at higher levels after skin injury, and one of these, heparin-binding EGF, has been shown to have a functional role in re-epithelialization<sup>48</sup> (Fig. 2). However, the large number of EGFR ligands at the wound site probably allows functional redundancy; therefore, ascertaining how the EGF family contributes to the wound repair process will require studies in mice in which the *Egfr* gene is knocked out just in keratinocytes.

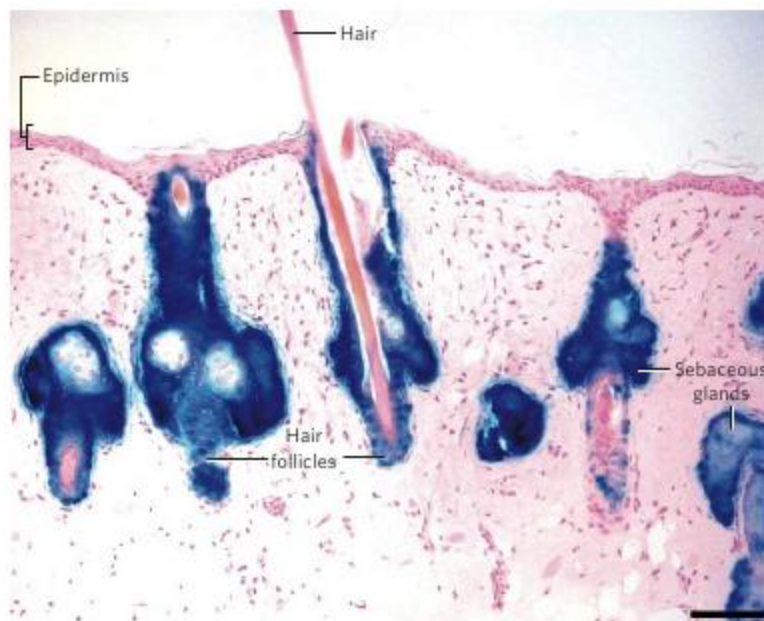
These positive regulators of re-epithelialization (that is, HGF, FGFs and EGFs) are ligands of receptor tyrosine kinases, the activation of which most often stimulates keratinocytes to migrate, proliferate and survive. Some of these processes are mediated by AP1-family members and by another transcription factor, STAT3. These factors are activated in response to signalling through various receptor tyrosine kinases, as well as in response to cytokine-receptor activation. Loss of AP1 proteins or AP1 components impaired dorsal closure and wound repair in *D. melanogaster*<sup>20,49</sup>. Recent studies suggest that AP1 proteins have a similar role in mammalian wound repair<sup>50</sup>, although some of their functions are probably masked by redundancy between the various members of the mammalian AP1 family<sup>49</sup>. STAT3, however, is clearly involved in re-epithelialization, because epidermis-specific deletion of *Stat3* in mice was found to retard epithelial repair after wounding<sup>51</sup>.

In contrast to these mitogenic growth factors, TGF- $\beta$  is a negative regulator of wound re-epithelialization. Re-epithelialization was shown to be strongly accelerated in mice expressing a gene encoding a dominant-negative TGF- $\beta$  receptor in the epidermis<sup>52</sup> and in mice lacking the transcriptional regulator SMAD3, one of the main targets of TGF- $\beta$ -mediated signalling<sup>53</sup>. Surprisingly, the TGF- $\beta$ -family member activin, which also signals through SMAD3, does not inhibit keratinocyte proliferation at the wound site but, instead, promotes proliferation. This effect, however, is probably mediated through the mesenchyme<sup>54</sup>, further highlighting the importance of epithelial-mesenchymal interactions in the wound repair process.

### Hormones and other factors

In addition to peptide growth factors, several low-molecular-mass mediators are regulators of wound re-epithelialization. For example, the hormone acetylcholine and its receptors are produced by keratinocytes, resulting in an autocrine loop that both positively regulates migration (through muscarinic acetylcholine receptors of the M4 subtype) and negatively regulates migration (through M3 receptors)<sup>55</sup>. Keratinocytes also produce hormones known as catecholamines (including adrenaline) and their receptors, resulting in the inhibition of re-epithelialization in an autocrine manner<sup>56</sup>.





**Figure 3 | Epithelial stem-cell-mediated skin regeneration.** To examine which cell types contribute to epidermal renewal, a single multipotent stem cell was isolated from a whisker follicle of a rat, labelled with a retrovirus carrying *lacZ* (which encodes  $\beta$ -galactosidase), expanded in culture and transplanted into newborn mouse skin<sup>61</sup>. The skin was biopsied 7 months later, and the tissue section shown indicates  $\beta$ -galactosidase activity (blue) and cell nuclei (pink). The progeny of the labelled stem cell contributed to several hair follicles and sebaceous glands (blue) but not to the epidermis; it should be noted that the follicles are in telogen (resting phase). Hence, hair-follicle stem cells do not contribute to long-term epidermal renewal, and the epidermis contains its own stem cells. Scale bar, 100  $\mu$ m. (Image courtesy of S. Claudinot, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland.)

Unexpectedly, polyunsaturated fatty acids and fatty-acid derivatives that activate peroxisome-proliferator-activated receptors (PPARs) have also been found to be important regulators of re-epithelialization. Expression of PPAR- $\alpha$  and PPAR- $\beta$  (also known as PPAR- $\delta$ ) is increased in keratinocytes after skin injury<sup>57</sup> (Fig. 2). Upregulation of expression of the gene encoding PPAR- $\beta$  and of as-yet unidentified ligands is achieved through the actions of pro-inflammatory cytokines, which in turn activate AP1 proteins through the stress-activated protein-kinase signalling cascade. This is functionally important, because mice in which the gene encoding PPAR- $\beta$  was knocked out showed a significantly reduced rate of re-epithelialization, and this defect resulted from reduced migration and increased apoptosis of keratinocytes<sup>58</sup>. PPAR- $\beta$  increases cell survival by upregulating expression of the genes encoding integrin-linked kinase and 3-phosphoinositide-dependent protein kinase 1, which in turn phosphorylate and activate the anti-apoptotic protein AKT<sup>59</sup>.

Finally, an unexpected recent finding is that electrical signals also regulate wound re-epithelialization<sup>60</sup>. It is well known that disruption of an epithelial layer generates an endogenous electric field in a lateral plane. This field was found to be an important directional cue for the migration of keratinocytes in response to wounding of a monolayer *in vitro*. The receptors and signalling molecules shown to be involved include EGFR,  $\alpha_5\beta_1$ -integrin, phosphatidylinositol-3-OH kinase- $\gamma$  and the phosphatase PTEN<sup>43,60</sup>. This mechanism regulates repair of the injured cornea and might have a similar role in wounded skin<sup>43,60</sup>.

It is clear, therefore, that a variety of factors, including physical factors, can activate the intracellular signalling pathways that, ultimately, regulate the various steps of wound re-epithelialization.

### Epithelial stem-cell biology

During re-epithelialization, renewal of the epidermis is required to provide keratinocytes, which then migrate and cover the healing wound. Similarly to other lining epithelia, such as those at the surface of the eye and the gut, the epidermis constantly and rapidly renews itself, a process

that allows homeostasis and the maintenance of proper tissue function. This involves the proliferation of epidermal stem cells. Stem cells are defined by their capacity to self-renew for an extended period of time (like cancer stem cells) and their ability to differentiate into mature, adult cells. Stem cells can be defined as pluripotent (capable of forming all the cell types of the body, including germ cells) or multipotent (capable of forming many cell types). As is the case in haematopoiesis, multipotent epithelial stem cells can be distinguished from multipotent progenitor cells only by a combination of clonal analysis and serial long-term transplantation experiments<sup>61</sup>. By using this approach, it has been confirmed unambiguously that the upper permanent region of the hair follicle below the sebaceous glands (commonly referred to as the bulge or the niche) contains multipotent progenitor cells<sup>62,63</sup> (Fig. 3). The same experimental approach has also indisputably shown the presence of multipotent stem cells outside the bulge, confirming previous results indicating that stem cells are not exclusively located in the bulge<sup>64</sup>.

Tissue stem cells can also broaden their capacity to form various lineages in response to physiological stimuli or injuries, a property that has great potential for regenerative medicine approaches. It has been argued that the epidermis is renewed by multipotent progenitor cells or stem cells generated in, and migrating from, the hair-follicle bulge<sup>62</sup>. Undoubtedly, multipotent stem cells from the bulge can contribute to epidermal repair<sup>61,65</sup>, but this occurs only when a wound cannot spontaneously repair itself through the migration of epidermal cells from the neighbouring unwounded epidermis or from the infundibulum, the portion of the hair follicle between the epidermis and the sebaceous gland (Fig. 3). Indeed, the infundibulum — which can extend deep into the dermis, up to several hundred micrometres in human hair follicles — contains distinct epidermal and hair-follicle territories (Fig. 3). Furthermore, genetic analyses have confirmed that the epidermis is self-renewing and that it does not depend on cells generated from multipotent stem cells of the hair follicle<sup>65,66</sup>. Finally, lineage-tracing studies in mice have revisited the classic concept that epidermal renewal is based on a hierarchy of stem cells and transient amplifying cells<sup>67</sup>, in support of the hypothesis that epidermal renewal relies on a single independent population of proliferative cells during normal homeostasis.

Another consideration is that the features of stem cells are still under debate, and this therefore influences experimental design and interpretation. Quiescence has been thought to be an important property of all stem cells, so retention of a label (indicating that cells are not actively dividing) has long been an indispensable criterion for the identification of epithelial stem cells<sup>68</sup>. However, there is now compelling evidence that stem cells can divide rapidly in some tissues but might not in others. Recently, it was shown that intestinal stem cells — as identified by expression of *Lgr5* (which encodes leucine-rich-repeat-containing G-protein-coupled receptor 5, a transmembrane protein downstream of the WNT-mediated signalling pathway) — are rapidly cycling<sup>69</sup>. Together with the observation that haematopoietic stem cells do not necessarily retain label<sup>70</sup>, these findings have led some researchers to reconsider quiescence as a key feature of 'stemness'; hence, the absence of label-retaining cells, as is the case for the expression of differentiation markers, does not preclude the presence of stem cells in a tissue. Ultimately, function is the only property that can properly identify stemness. This could explain why mouse corneas can self-renew in the complete absence of limbus, the transitional zone at the junction of the conjunctiva and the cornea, which is thought to contain the corneal stem cells<sup>71</sup>. The identification of all cells that can function as stem cells is certain to influence future strategies for wound repair and tissue regeneration.

Adult corneal cells exposed to embryonic dermis can be induced to form hair follicles<sup>72</sup>. This finding, together with the observation that NOTCH1-deficient corneal epithelium forms an epidermis when wounded<sup>71</sup>, indicates that epithelial cells have some plasticity, a property that has important clinical implications. Indeed, cultured grafts of epithelium obtained from the oral cavity have been used to reconstruct human corneas<sup>73</sup>. Advances in stem-cell biology have implications for improving skin therapies outside the traditional domain of wound repair. For example, the recent finding that stem cells from an individual



can be genetically engineered and then permanently engrafted in the individual for the treatment of incapacitating hereditary skin diseases is a considerable step forward<sup>74</sup>. It not only shows that *ex vivo* gene therapy is feasible but also brings new hope to patients with chronic wounds for which conventional wound repair therapy has failed. Despite such successes, there is much to accomplish before stem cells can be safely manipulated to reconstruct the function of the skin and other stratified epithelia fully.

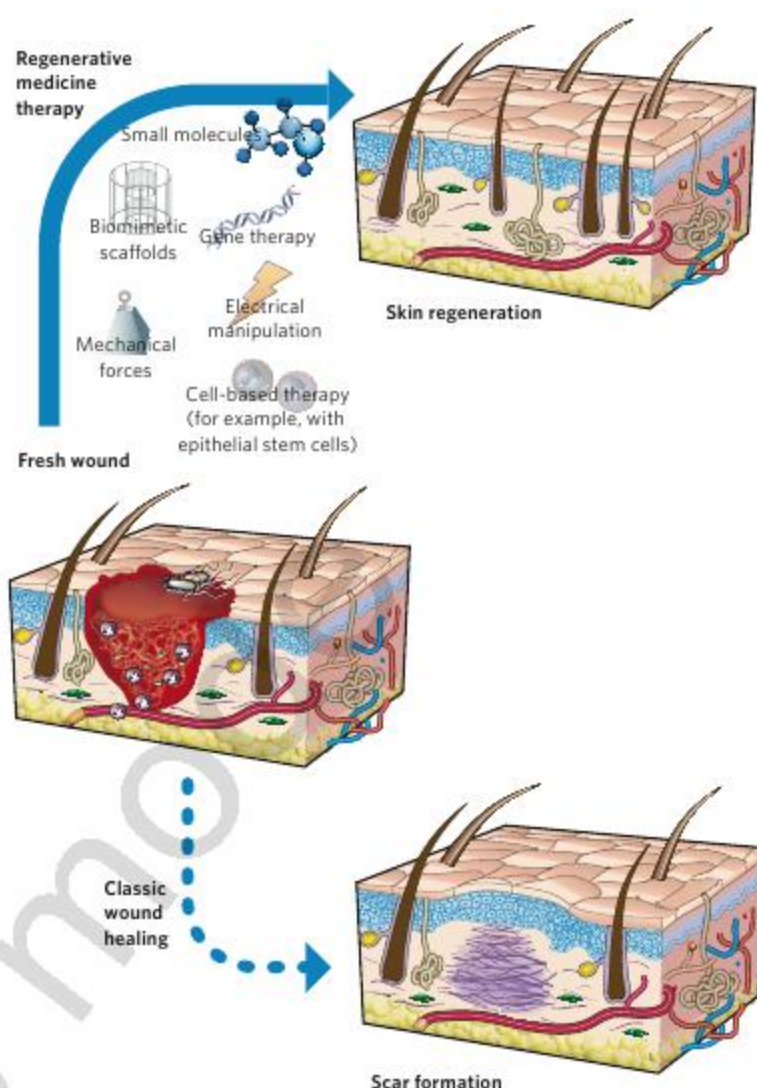
In addition, the capacity to massively expand multipotent epithelial stem cells in culture, together with a better understanding of the epithelial–mesenchymal interactions that control hair-follicle morphogenesis<sup>75,76</sup>, opens the door to the *ex vivo* reconstruction of skin appendages. In this regard, a recent report of spontaneous generation of hair follicles in the healing skin of adult mice emphasizes that WNT-mediated signalling is important for skin morphogenesis and repair<sup>77</sup> and provides hope that hair follicles can be newly generated. But *de novo* hair-follicle morphogenesis, which is known to occur in some species (for example, in deer during antler growth), has never been observed in humans. Of equal importance is the inductive role of dermal papilla cells in hair-follicle genesis and their relationship to the neural crest<sup>78–80</sup>. Dermal papilla cells can express markers of adipogenic, chondrogenic and osteogenic differentiation and can even form neurosphere-like spheres (which express markers of neuronal differentiation when cultured in the appropriate conditions). Therefore, dermal papilla cells seem to be functionally related to mesenchymal cells isolated from other tissues (for example, the adipose tissue or the bone marrow). Understanding the molecular control of epithelial and mesenchymal stem-cell fate will allow the design of new strategies for wound repair: for example, strategies to enhance the migration of stem cells, to induce the formation of epidermal appendages (such as hair follicles and sweat duct glands) and to decrease scarring<sup>75,76,81</sup>.

### Towards tissue regeneration in humans

In humans, problems with wound healing can manifest as either delayed wound healing (which occurs with diabetes or radiation exposure) or excessive healing (as occurs with hypertrophic and keloid scars). Excessive healing is characterized by the deposition of large amounts of extracellular matrix and by alterations in local vascularization and cell proliferation. These excessive fibrotic reactions manifest in humans as a 'bad scar'. Many such instances of 'overhealing' produce large disfiguring masses that can physically distort surface structures (such as the nose or eyelids). These commonly occur after major injuries such as burns, in which case they are referred to as hypertrophic scars. They can also appear for unknown reasons after a relatively minor trauma, as is the case for keloid scars, which might have a genetic basis<sup>2</sup>. Much interest has been generated by the observation that increased amounts of TGF- $\beta$  are found in wounds that heal by scar formation as opposed to tissue regeneration as seen in bad scars<sup>82,83</sup>. This finding has led to clinical efforts to block scar formation with antibodies and small molecules directed against TGF- $\beta$  and other pro-inflammatory mediators<sup>84</sup>. Recent evidence also suggests that changes in the physical environment might result in overhealing, by affecting the mechanical environment of the cells in the wound<sup>85</sup>.

Although tremendous strides have been made in delineating the myriad factors involved in normal and pathological tissue repair, these findings have not led to substantial advances in patient care. It has become clear that single-agent therapies, such as administration of a growth factor, have only a moderate impact on wound repair in a clinical setting, most probably because of the considerable plasticity and redundancy of the components of the wound repair process or because of their rapid degradation at the wound site.

The ultimate solution to both underhealing and overhealing is likely to be administration of cells that retain the ability to elaborate the full complexity of biological signalling, together with the environmental cues that are needed to regulate the differentiation and proliferation of these cells. Previous attempts have used part of this approach (for example, the administration of cultured keratinocytes or fibroblasts), but until



**Figure 4 | Potential therapies for reducing scar formation during wound repair.** To manipulate wound repair to become more regenerative than scar forming, strategies include the use of biomimetic scaffolds, the manipulation of the mechanical environment (for example, negative-pressure wound therapy to increase healing) or the electrical environment, the administration of small molecules, the use of gene-therapy approaches, and the use of cell-based strategies (including administration of epithelial stem cells). All of these elements have been demonstrated to have an effect on *in vitro* and *in vivo* models of wound healing as single-agent therapies. In theory, many of these elements could be combined to recreate a receptive environment (or 'soil') to promote regeneration. Combining these with the appropriate stem cells (or 'seed') will undoubtedly alter the result of wound healing in humans.

now it has not been possible to provide both 'seed and soil' in the same therapeutic agent. Recent advances in stem-cell and progenitor-cell biology have resulted in the isolation and characterization of skin progenitor cells in mammals. In parallel, advances in material science have made it possible to deliver extracellular or intracellular signals precisely in the appropriate temporal and spatial sequence<sup>86</sup>. One such approach is illustrated in Fig. 4. Adult epidermal progenitor cells (obtained from skin samples or standardized cell lines) are placed into a biomimetic matrix that reproduces the environment present during prenatal growth and development, when tissue regeneration occurs. Components of the environment include optimized mechanical stress and oxygen tension, and extracellular-matrix proteins. After grafting to the injured site, these progenitor cells can themselves provide the proper sequence of extracellular factors to accelerate tissue repair and regeneration. Using cells to integrate environmental signals and transduce them into biological effectors is likely to be crucial for new approaches to the ubiquitous problem of fibrotic wound repair. Thus, researchers and doctors stand at the dawn of a new era in which cell function can be controlled and regulated in clinical situations.



1. Singer, A. J. & Clark, R. A. Cutaneous wound healing. *N. Engl. J. Med.* **341**, 738–746 (1999).
2. Aarabi, S., Longaker, M. T. & Gurtner, G. C. Hypertrophic scar formation following burns and trauma: new approaches to treatment. *PLoS Med.* **4**, e234 (2007).
3. Konigova, R. & Rychterova, V. Marjolin's ulcer. *Acta Chir. Plast.* **42**, 91–94 (2000).
4. Trent, J. T. & Kirsner, R. S. Wounds and malignancy. *Adv. Skin Wound Care* **16**, 31–34 (2003).
5. Colwell, A. S., Longaker, M. T. & Lorenz, H. P. Fetal wound healing. *Front. Biosci.* **8**, s1240–s1248 (2003).
6. Gurtner, G. C., Callaghan, M. J. & Longaker, M. T. Progress and potential for regenerative medicine. *Annu. Rev. Med.* **58**, 299–312 (2007).
7. National Heart, Lung, and Blood Institute. *Morbidity & Mortality: 2002 Chart Book on Cardiovascular, Lung, and Blood Diseases* (US Department of Health and Human Services, Bethesda, 2002).
8. Anderson, R. N. & Smith, B. L. Deaths: leading causes for 2001. *Natl. Vital Stat. Rep.* **52** (9), 1–85 (2003).
9. Selman, M., King, T. E. & Pardo, A. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Ann. Intern. Med.* **134**, 136–151 (2001).
10. Mescher, A. L. & Neff, A. W. Regenerative capacity and the developing immune system. *Adv. Biochem. Eng. Biotechnol.* **93**, 39–66 (2005).
11. Klein, L. et al. Pharmacologic therapy for patients with chronic heart failure and reduced systolic function: review of trials and practical considerations. *Am. J. Cardiol.* **91**, 18F–40F (2003).
12. Nichols, S. A., Dirks, W., Pearce, J. S. & King, N. Early evolution of animal cell signaling and adhesion genes. *Proc. Natl. Acad. Sci. USA* **103**, 12451–12456 (2006).
13. Adamska, M. et al. The evolutionary origin of hedgehog proteins. *Curr. Biol.* **17**, R836–R837 (2007).
14. Adamska, M. et al. Wnt and TGF- $\beta$  expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. *PLoS ONE* **2**, e1031 (2007).
15. Adamska, M. et al. The evolutionary origin of hedgehog proteins. *Curr. Biol.* **17**, R836–R837 (2007).
16. Woolner, S., Jacinto, A. & Martin, P. The small GTPase Rac plays multiple roles in epithelial sheet fusion — dynamic studies of *Drosophila* dorsal closure. *Dev. Biol.* **282**, 163–173 (2005).
17. Samakovlis, C. et al. Genetic control of epithelial tube fusion during *Drosophila* tracheal development. *Development* **122**, 3531–3536 (1996).
18. Wood, W. et al. Wound healing recapitulates morphogenesis in *Drosophila* embryos. *Nature Cell Biol.* **4**, 907–912 (2002).
19. Woolley, K. & Martin, P. Conserved mechanisms of repair: from damaged single cells to wounds in multicellular tissues. *Bioessays* **22**, 911–919 (2000).
20. Martin, P. & Parkhurst, S. M. Parallels between tissue repair and embryo morphogenesis. *Development* **131**, 3021–3034 (2004).
21. Suzuki, M., Satoh, A., Ide, H. & Tamura, K. Nerve-dependent and -independent events in blastema formation during *Xenopus* froglet limb regeneration. *Dev. Biol.* **286**, 361–375 (2005).
22. Brockes, J. P. & Kumar, A. Plasticity and reprogramming of differentiated cells in amphibian regeneration. *Nature Rev. Mol. Cell Biol.* **3**, 566–574 (2002).
23. Endo, T., Bryant, S. V. & Gardiner, D. M. A stepwise model system for limb regeneration. *Dev. Biol.* **270**, 135–145 (2004).
24. Kumar, A., Godwin, J. W., Gates, P. B., Garza-Garcia, A. A. & Brockes, J. P. Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. *Science* **318**, 772–777 (2007).
25. Sánchez Alvarado, A. Planarian regeneration: its end is its beginning. *Cell* **124**, 241–245 (2006).
26. Klapka, N. & Müller, H. W. Collagen matrix in spinal cord injury. *J. Neurotrauma* **23**, 422 (2006).
27. Stichel, C. C. & Müller, H. W. The CNS lesion scar: new vistas on an old regeneration barrier. *Cell Tissue Res.* **294**, 1–9 (1998).
28. Grose, R. & Werner, S. Wound-healing studies in transgenic and knockout mice. *Mol. Biotechnol.* **28**, 147–166 (2004).
29. Martin, P. & Leibovich, S. J. Inflammatory cells during wound repair: the good, the bad and the ugly. *Trends Cell Biol.* **15**, 599–607 (2005).
30. Martin, P. et al. Wound healing in the PU.1 null mouse — tissue repair is not dependent on inflammatory cells. *Curr. Biol.* **13**, 1122–1128 (2003).
31. Werner, S. & Grose, R. Regulation of wound healing by growth factors and cytokines. *Physiol. Rev.* **83**, 835–870 (2003).
32. Galiano, R. D. et al. Topical vascular endothelial growth factor accelerates diabetic wound healing through increased angiogenesis and by mobilizing and recruiting bone marrow-derived cells. *Am. J. Pathol.* **164**, 1935–1947 (2004).
33. Bluff, J. E., Ferguson, M. W. J., O'Kane, S. & Ireland, G. Bone marrow-derived endothelial progenitor cells do not contribute significantly to new vessels during incisional wound healing. *Exp. Hematol.* **35**, 500–506 (2007).
34. Opalenik, S. R. & Davidson, J. M. Fibroblast differentiation of bone marrow-derived cells during wound repair. *FASEB J.* **19**, 1561–1563 (2005).
35. Werner, S., Krieg, T. & Smola, H. Keratinocyte-fibroblast interactions in wound healing. *J. Invest. Dermatol.* **127**, 998–1008 (2007).
36. Szabowski, A. et al. c-Jun and JunB antagonistically control cytokine-regulated mesenchymal-epidermal interaction in skin. *Cell* **103**, 745–755 (2000).
37. Lovvorn, H. N. et al. Relative distribution and crosslinking of collagen distinguish fetal from adult sheep wound repair. *J. Pediatr. Surg.* **34**, 218–223 (1999).
38. Levenson, S. M. et al. The healing of rat skin wounds. *Ann. Surg.* **161**, 293–308 (1965).
39. Rubin, G. M. et al. Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
40. Cole, J., Tsou, R., Wallace, K., Gibrán, N. & Isik, F. Early gene expression profile of human skin to injury using high-density cDNA microarrays. *Wound Repair Regen.* **9**, 360–370 (2001).
41. Cooper, L., Johnson, C., Burslem, F. & Martin, P. Wound healing and inflammation genes revealed by array analysis of 'macrophageless' PU.1 null mice. *Genome Biol.* **6**, R5 (2005).
42. Chang, H. Y. et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, E7 (2004).
43. Raja, K., Sivamani, M. S., Garcia, R. R. & Isseroff, R. R. Wound re-epithelialization: modulating keratinocyte migration in wound healing. *Front. Biosci.* **12**, 2849–2868 (2007).
44. Chmielowiec, J. et al. c-Met is essential for wound healing in the skin. *J. Cell Biol.* **177**, 151–162 (2007).
45. Werner, S. et al. The function of KGF in morphogenesis of epithelium and reepithelialization of wounds. *Science* **266**, 819–822 (1994).
46. Braun, S., auf dem Keller, U., Steiling, H. & Werner, S. Fibroblast growth factors in epithelial repair and cytoprotection. *Phil. Trans. R. Soc. Lond. B* **359**, 753–757 (2004).
47. Jameson, J. et al. A role for skin  $\gamma\delta$  T cells in wound repair. *Science* **296**, 747–749 (2002).
48. Shirakata, Y. et al. Heparin-binding EGF-like growth factor accelerates keratinocyte migration and skin wound healing. *J. Cell Sci.* **118**, 2363–2370 (2005).
49. Schäfer, M. & Werner, S. Transcriptional control of wound repair. *Annu. Rev. Cell Dev. Biol.* **23**, 69–92 (2007).
50. Li, G. et al. c-Jun is essential for organization of the epidermal leading edge. *Dev. Cell* **4**, 865–877 (2003).
51. Sano, S. et al. Keratinocyte-specific ablation of Stat3 exhibits impaired skin remodeling, but does not affect skin morphogenesis. *EMBO J.* **18**, 4657–4668 (1999).
52. Amendt, C., Mann, A., Schirmacher, P. & Blessing, M. Resistance of keratinocytes to TGF- $\beta$ -mediated growth restriction and apoptosis induction accelerates re-epithelialization in skin wounds. *J. Cell Sci.* **115**, 2189–2198 (2002).
53. Ashcroft, G. S. et al. Mice lacking Smad3 show accelerated wound healing and an impaired local inflammatory response. *Nature Cell Biol.* **1**, 260–266 (1999).
54. Werner, S. & Alzheim, C. Roles of activin in tissue repair, fibrosis, and inflammatory disease. *Cytokine Growth Factor Rev.* **17**, 157–171 (2006).
55. Chernyavsky, A. I., Arredondo, J., Wess, J. R., Karlsson, E. & Grando, S. A. Novel signaling pathways mediating reciprocal control of keratinocyte migration and wound epithelialization through M3 and M4 muscarinic receptors. *J. Cell Biol.* **166**, 261–272 (2004).
56. Pullar, C. E., Rizzo, A. & Isseroff, R. R.  $\beta$ -Adrenergic receptor antagonists accelerate skin wound healing: evidence for a catecholamine synthesis network in the epidermis. *J. Biol. Chem.* **281**, 21225–21235 (2006).
57. Michalik, L. et al. Impaired skin wound healing in peroxisome proliferator-activated receptor (PPAR) $\alpha$  and PPAR $\beta$  mutant mice. *J. Cell Biol.* **154**, 799–814 (2001).
58. Icre, G., Wahli, W. & Michalik, L. Functions of the peroxisome proliferator-activated receptor (PPAR) $\alpha$  and  $\beta$  in skin homeostasis, epithelial repair, and morphogenesis. *J. Invest. Dermatol. Symp. Proc.* **11**, 30–35 (2006).
59. Di-Pol, N., Tan, N. S., Michalik, L., Wahli, W. & Desvergne, B. Antiapoptotic role of PPAR $\beta$  in keratinocytes via transcriptional control of the Akt1 signaling pathway. *Mol. Cell* **10**, 721–733 (2002).
60. Zhao, M. et al. Electrical signals control wound healing through phosphatidylinositol-3-OH kinase- $\gamma$  and PTEN. *Nature* **442**, 457–460 (2006).
61. Claudinot, S. P., Nicolas, M., Oshima, H., Rochat, A. & Barrandon, Y. Long-term renewal of hair follicles from clonogenic multipotent stem cells. *Proc. Natl. Acad. Sci. USA* **102**, 14677–14682 (2005).
62. Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102**, 451–461 (2000).
63. Oshima, H., Rochat, A., Kedzia, C., Kobayashi, K. & Barrandon, Y. Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell* **104**, 233–245 (2001).
64. Rochat, A., Kobayashi, K. & Barrandon, Y. Location of stem cells of human hair follicles by clonal analysis. *Cell* **76**, 1063–1073 (1994).
65. Ito, M. et al. Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11**, 1351–1354 (2005).
66. Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev. Cell* **9**, 855–861 (2005).
67. Clayton, E. et al. A single type of progenitor cell maintains normal epidermis. *Nature* **446**, 185–189 (2007).
68. Potten, C. S. & Booth, C. Keratinocyte stem cells: a commentary. *J. Invest. Dermatol.* **119**, 888–899 (2002).
69. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
70. Kiel, M. J. et al. Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. *Nature* **449**, 238–242 (2007).
71. Vauclair, S. et al. Corneal epithelial cell fate is maintained during repair by Notch1 signaling via the regulation of vitamin A metabolism. *Dev. Cell* **13**, 242–253 (2007).
72. Pearton, D. J., Yang, Y. & Dhoubilly, D. Transdifferentiation of corneal epithelium into epidermis occurs by means of a multistep process triggered by dermal developmental signals. *Proc. Natl. Acad. Sci. USA* **102**, 3714–3719 (2005).
73. Nishida, K. et al. Corneal reconstruction with tissue-engineered cell sheets composed of autologous oral mucosal epithelium. *N. Engl. J. Med.* **351**, 1187–1196 (2004).
74. Mavilio, F. et al. Correction of junctional epidermolysis bullosa by transplantation of genetically modified epidermal stem cells. *Nature Med.* **12**, 1397–1402 (2006).



75. Metcalfe, F. & Ferguson, F. Skin stem and progenitor cells: using regeneration as a tissue-engineering strategy. *Cell. Mol. Life Sci.* **65**, 24–32 (2008).
76. Metcalfe, A. D. & Ferguson, M. W. J. Tissue engineering of replacement skin: the crossroads of biomaterials, wound healing, embryonic development, stem cells and regeneration. *J. R. Soc. Interface* **4**, 413–437 (2007).
77. Ito, M. *et al.* Wnt-dependent *de novo* hair follicle regeneration in adult mouse skin after wounding. *Nature* **447**, 316–320 (2007).
78. Reynolds, A. J., Lawrence, C., Cserhalmi-Friedman, P. B., Christiano, A. M. & Jahoda, C. A. Trans-gender induction of hair follicles. *Nature* **402**, 33–34 (1999).
79. Fernandes, K. J. L. *et al.* A dermal niche for multipotent adult skin-derived precursor cells. *Nature Cell Biol.* **6**, 1082–1093 (2004).
80. Wong, C. E. *et al.* Neural crest-derived cells with stem cell features can be traced back to multiple lineages in the adult skin. *J. Cell Biol.* **175**, 1005–1015 (2006).
81. Blanpain, C., Lowry, W. E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635–648 (2004).
82. Lin, R. Y. *et al.* Exogenous transforming growth factor- $\beta$  amplifies its own expression and induces scar formation in a model of human fetal skin repair. *Ann. Surg.* **222**, 146–154 (1995).
83. Ghahary, A., Shen, Y. J., Scott, P. G., Gong, Y. & Tredget, E. E. Enhanced expression of mRNA for transforming growth factor- $\beta$ , type I and type III procollagen in human post-burn hypertrophic scar tissues. *J. Lab. Clin. Med.* **122**, 465–473 (1993).
84. Ferguson, M. W. & O'Kane, S. Scar-free healing: from embryonic mechanisms to adult therapeutic intervention. *Phil. Trans. R. Soc. Lond. B* **359**, 839–850 (2004).
85. Arabi, S. *et al.* Mechanical load initiates hypertrophic scar formation through decreased cellular apoptosis. *FASEB J.* **21**, 3250–3261 (2007).
86. Lutolf, M. P. & Hubbell, J. A. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. *Nature Biotechnol.* **23**, 47–55 (2005).

This review comprehensively describes state-of-the-art synthetic biomaterials for guiding cell differentiation and tissue regeneration.

**Acknowledgements** G.C.G. and M.T.L. are supported by grants from the National Institutes of Health and the Oak Foundation. S.W. is supported by the Swiss Federal Institute of Technology Zürich (ETHZ), the Swiss National Science Foundation (SNSF), the European Union and Oncosuisse. Y.B. is supported by the École Polytechnique Fédérale de Lausanne (EPFL), the Centre Hospitalier Universitaire Vaudois Lausanne (CHUV) and the European Consortium for Stem Cell Research (EuroStemCell).

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to G.C.G. or M.T.L. ([ggurtner@stanford.edu](mailto:ggurtner@stanford.edu); [longaker@stanford.edu](mailto:longaker@stanford.edu)).



# Stem-cell-based therapy and lessons from the heart

Robert Passier<sup>1,2</sup>, Linda W. van Laake<sup>1,3</sup> & Christine L. Mummery<sup>1,2</sup>

**The potential usefulness of human embryonic stem cells for therapy derives from their ability to form any cell in the body. This potential has been used to justify intensive research despite some ethical concerns. In parallel, scientists have searched for adult stem cells that can be used as an alternative to embryonic cells, and, for the heart at least, these efforts have led to promising results. However, most adult cardiomyocytes are unable to divide and form new cardiomyocytes and would therefore be unable to replace those lost as a result of disease. Basic questions — for example, whether cardiomyocyte replacement or alternatives, such as providing the damaged heart with new blood vessels or growth factors to activate resident stem cells, are the best approach — remain to be fully addressed. Despite this, preclinical studies on cardiomyocyte transplantation in animals and the first clinical trials with adult stem cells have recently been published with mixed results.**

Embryonic stem (ES) cells from mice and primates can differentiate into any cell type in the adult body<sup>1</sup>. The potential for using differentiated human ES cells to treat degenerative disease and injuries has long been evident. However, the ethical debate that ensued from human ES cell derivation requiring the destruction of human embryos caused antipathy in many countries, and it prompted huge efforts to find alternative approaches that could use adult stem cells or progenitor cells instead. In terms of investment, the National Institutes of Health spent US\$2.5 billion on stem-cell research in the period 2004–2007, of which only \$150 million was on human ES-cell research (<http://www.nih.gov/news/fundingresearchareas.htm>).

It has been assumed that if enough correct cells can be generated, they would be able to replace cells that have been lost or damaged by disease. This concept is particularly pertinent for the heart, where new heart muscle cells (cardiomyocytes) should be able to restore contractile function after a heart attack (myocardial infarction). It would also be useful for the pancreas, where new  $\beta$ -cells could secrete insulin in response to glucose after transplantation into people with diabetes, and in the central nervous system, where transplantation of appropriate neurons into the brain could restore dopamine production in patients with Parkinson's disease. In addition, transplanting stem cells into the spinal cord would allow new neurons to restore motor function in people with spinal-cord injuries.

Independently of the ethical issues, however, human ES-cell research has encountered many obstacles on its route to the clinic. These include the risk of forming a teratoma, a benign tumour containing mixtures of disorganized, but identifiable, tissues such as cartilage and bone. Teratomas are caused by the presence of contaminating undifferentiated cells if they are inadvertently co-transplanted with differentiated cells. Another risk is an immune response, which would destroy transplanted cells not matched with the recipient. Other obstacles to be overcome are ensuring the clinical compliance of culture reagents so that they do not, for example, contain animal reagents possibly contaminated with viruses or prions and, as has emerged more recently with publication of preclinical transplantation studies in disease models in animals, problems of cell survival in grafts and difficulties with proper integration into the

host tissue. The risk of teratoma formation is specific to human ES cells and human induced pluripotent stem (iPS) cells, which were recently derived by direct reprogramming of skin fibroblasts to a pluripotent state (in which they can give rise to all the cell types of the body) (Box 1), whereas postnatal or adult stem cells could in principle be used directly, without predifferentiation or culture, although poor integration and survival in host tissues would still be a problem. Adult stem cells could also be used autologously. (These stem cells are from the patient's body and therefore there would be no immune rejection after transplantation.) If derived from a stem-cell-rich source such as the bone marrow or cord blood, they may not need expansion in culture before use, thus avoiding exposure to animal-derived reagents and preventing the development of chromosomal abnormalities<sup>2</sup>, which can occur if cells become stressed. These advantages have led to rapid clinical application of autologous adult stem cells.

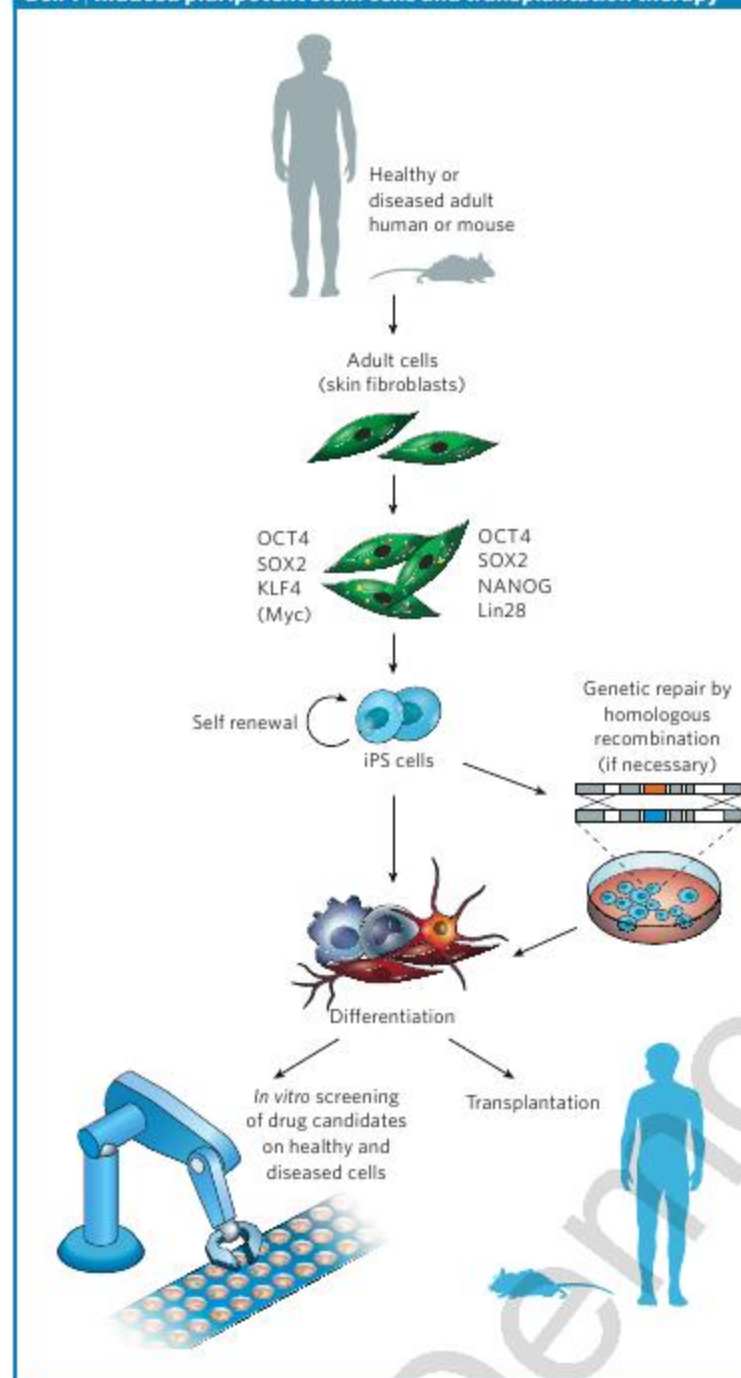
Here, we review the status of research on cell transplantation in experimental animals and humans. We focus on stem cells that have been tested for cardiac therapy. These cells have been largely derived from haematopoietic sources such as the bone marrow and the blood because they are arguably the most advanced in terms of transplantation into animals and clinical trials. We also consider cells that have been isolated and cultured from the adult heart; these have the properties of fetal cardiac progenitor cells, even though their direct role in heart regeneration *in vivo* has yet to be unequivocally established (Fig. 1). We also discuss the current state of transplantation of cardiomyocytes derived from human ES cells in clinically relevant animal models for cardiac disease, because the challenges are relevant to transplantation of cardiomyocytes from any stem-cell source to the adult heart and, indeed, of transplantation of differentiated cells into a variety of organs and tissues.

## Stem cells in fetal and adult hearts

The heart has limited intrinsic regenerative capacity. Damaged and diseased cardiomyocytes are removed largely by macrophages and replaced by scar tissue surrounded by poorly contractile survivors. Myocardial

<sup>1</sup>Hubrecht Institute, Developmental Biology and Stem Cell Research, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. <sup>2</sup>Department of Anatomy and Embryology, Leiden University Medical Centre, Einthovenweg 20, 2333 Leiden, The Netherlands. <sup>3</sup>University Medical Center Utrecht, Division Heart and Lungs, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.



**Box 1 | Induced pluripotent stem cells and transplantation therapy**

In 2006, it was shown, by Kazutoshi Takahashi and Shinya Yamanaka<sup>83</sup>, for the first time that overexpression of only four proteins — OCT4, SOX2, Myc and KLF4 — in mouse embryonic or adult fibroblast cultures could generate pluripotent stem-cell-like cells (see figure). Just a year later, the same group<sup>84</sup> and another group, led by James Thomson<sup>85</sup>, succeeded independently in generating human induced pluripotent stem (iPS) cells from adult skin fibroblasts. Thomson's group used overexpression of OCT4 and SOX2, in combination with two other proteins, NANOG and Lin28, instead of Myc and KLF4. In both studies, iPS cells showed the essential characteristics of ES cells in terms of their morphology, cell-surface markers, gene-expression profiles and telomerase activity. Furthermore, iPS cell clones could be maintained in culture for several months at least and could be induced to differentiate into derivatives of all three embryonic germ layers both *in vitro* and *in vivo* in mouse teratomas. Reactivation of Myc increased tumorigenicity in chimaeric mice derived from mouse iPS cells, and a modified protocol was developed that did not require activation of Myc in either mouse or human cells<sup>86</sup>. Thus, it became feasible to generate iPS cells from fibroblast cultures from patients (with genetic defects corrected if necessary), and these cells could then, in principle, be induced to differentiate into a variety of patient-specific cell types, allowing transplantation without the risk of immune rejection (see figure).

Recently, Jacob Hanna and colleagues<sup>87</sup> provided a proof of principle for iPS-cell-based treatment in combination with genetic repair in a mouse model for sickle-cell anaemia, a genetic blood disorder. Tail-tip fibroblasts from these mice were infected with OCT4-, SOX2-, KLF4- and Myc-expressing viruses to generate iPS cells. (Myc copies were then deleted by using the recombinase Cre.) Subsequent homologous recombination in the iPS cells and correction of the genetic defect by the wild-type human variant rescued the phenotype; haematopoietic progenitor cells that were derived from these cells formed aggregates (known as embryoid bodies) in culture and were able to rescue the anaemic mice after transplantation, and the mice were cured. This experiment was the perfect combination of cell and gene therapy.

These studies open up exciting prospects for the treatment of various genetic diseases and other abnormalities that might benefit from cell-based therapy. However, the present methods for generating iPS cells require genetic integration by retroviruses or lentiviruses, even though oncogenes, such as Myc, no longer seem to be necessary<sup>86</sup>. Furthermore, in terms of transplantation, the issues of cell integration, survival and safety still need to be addressed. The more immediate applications of human iPS cells are likely to be the creation of human models of human disease *in vitro* for studying the underlying molecular mechanisms of disease, for screening drug candidates, and for assessing drug safety and toxicity (see figure).

infarction and other substantial or long-term injuries can lead to heart failure, which is fatal within 5 years for 65% of patients. It has long been assumed that the reason for poor regeneration of the adult heart was the absence of endogenous progenitor cells. Over the past few years, however, cell populations expressing stem-cell marker proteins such as Kit, stem-cell antigen 1 (SCA1) and multidrug resistance protein 1 (MDR1) have been identified in the human and/or mouse heart (Box 2), albeit in minuscule quantities (see ref. 3 for a review).

Although the initial evidence for an adult stem-cell or progenitor-cell population in the adult heart, which could potentially be harnessed for cardiac repair, was initially welcomed with enthusiasm, scepticism has since grown. These Kit-expressing cells in tissues of solid organs, including the heart, are thought to have left the bone marrow in minuscule quantities to scavenge pathogenic molecules in peripheral tissues as part of a mechanism to promote a local innate immune response<sup>4</sup>. They are not then actual heart cells but bone-marrow cells out of place. In addition, many of the cells expressing Kit that were detected in biopsied samples of adult human heart were recently reported to co-express markers of mast cells (cells of the immune system) and to lack expression of cardiac transcription factors NKX2-5 and islet 1 (Isl1)<sup>5</sup>, crucial markers of the cardiac progenitor cell state in fetal hearts<sup>6,7</sup>. These cells are therefore not

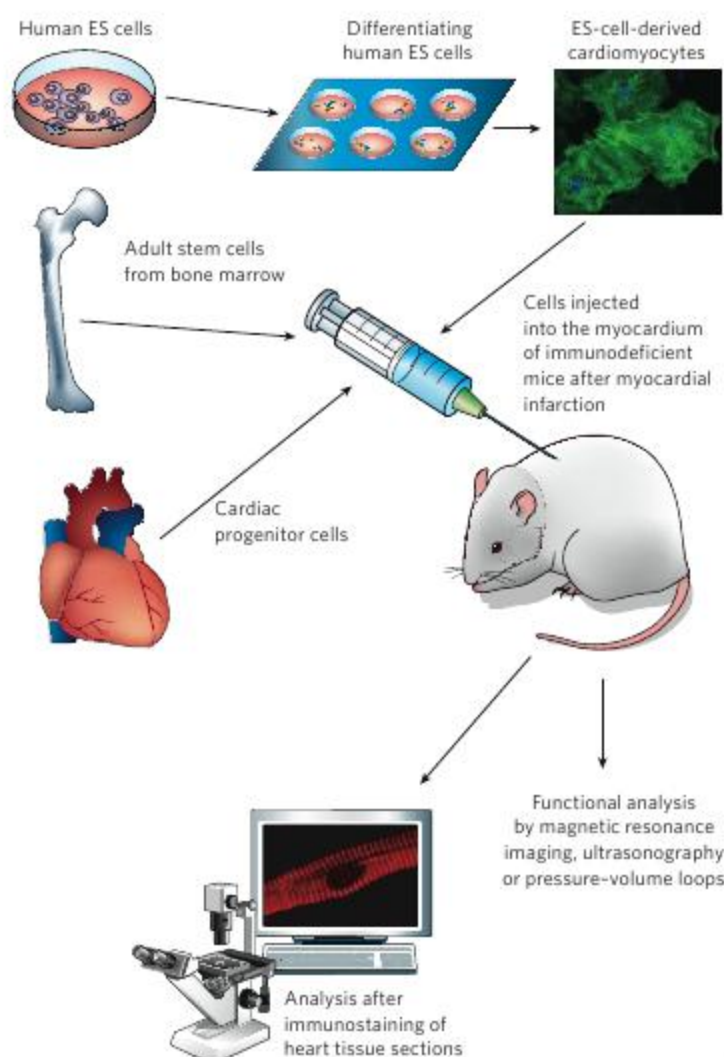
cardiac progenitor cells at all. *In vivo* lineage tracing in mice will be essential to resolve issues of the origin, identity and fate of these Kit-expressing cells as well as others expressing stem-cell marker proteins in the adult heart and their ability to contribute to endogenous repair.

### Adult stem cells and transplantation into the heart

The idea that certain stem cells can transdifferentiate into cell types outside their normal lineage has been raised only recently. In these reports, genetically marked postnatal cells injected into non-transgenic host animals were able to acquire the expression of tissue-specific markers<sup>8</sup>. The first apparent transdifferentiation of bone-marrow cells into cardiomyocytes was described shortly thereafter. Injection of green fluorescent protein (GFP)-marked bone-marrow cells from donor mice into the hearts of wild-type mice resulted in co-expression of cardiac proteins and restored heart function in mice that had suffered a myocardial infarction<sup>9</sup>. Given the unmet clinical need for therapy for patients with myocardial infarction and heart failure, this technique was quickly translated for use in patients, and the first randomized clinical trials were completed within 5 years.

As more reports on the outcome of bone-marrow-cell trials become available (see ref. 10 for a review), it is becoming unclear whether this





**Figure 1 | Transplantation strategy for cardiac repair.** Adult stem cells, cardiac progenitor cells or human ES-cell-derived cardiomyocytes are isolated and injected into the heart of immunodeficient mice that have had a myocardial infarction. At different time points after transplantation, cardiac function is analysed by using magnetic resonance imaging, ultrasonography or ventricular pressure-volume loops. To determine cell survival, phenotype and integration, hearts are isolated after transplantation, and transverse heart sections are used for immunostaining with specific fluorochrome-conjugated antibodies (cardiomyocytes are identified by antibodies specific for  $\alpha$ -actinin), followed by confocal microscopy.

approach provides any significant benefit. Although some early non-controlled pilot studies were unanimously positive with respect to improved cardiac function, the outcomes of recent randomized controlled trials have failed to produce the same outcome as in patients in non-controlled, non-randomized studies, especially after longer follow-up times. Patients who received injections of various numbers of cells from different cellular fractions of their own bone marrow had either no increase in the ejection fraction (the fraction of blood within the (left) ventricle ejected during one contraction) or, if they did, an increase that did not exceed 5% (refs 11–14), the minimum change required for long-term improvement of symptoms and survival<sup>15,16</sup>. By contrast, parameters such as infarct remodelling<sup>16</sup> and exercise capacity<sup>17</sup>, which may be more indicative of long-term outcome, were positively affected by bone-marrow-cell treatment, at least for the first 4–6 months. Patients with the largest infarcts generally benefited the most<sup>14,18</sup>.

Efforts to repeat the earlier studies that involved injecting bone-marrow cells into mice with a myocardial infarction failed to confirm transdifferentiation as the mechanism underlying the apparent expression of cardiac markers by bone-marrow cells, but they indicated, instead, that bone-marrow cells could fuse at a low frequency with host cardiomyocytes and express both sets of markers<sup>19–21</sup>. Whether

transdifferentiation takes place in humans is more difficult to investigate, but it would seem unlikely. The original results may also have been influenced by autofluorescence from scar tissue and/or dead cells<sup>22</sup>. The consensus now is that any improvement in cardiac function after bone-marrow-cell transplantation is likely to be the result of preservation of ischaemic myocardium (myocardium deprived of oxygen) by paracrine action from the transplanted cells, as opposed to an increase in the number of contractile cells by replacing those lost after damage. This hypothesis, that transplanted cells secrete factors beneficial for myocyte survival and blood-vessel formation, is supported by evidence from studies in rodents and larger mammals showing induction of neovascularization and rescue of ischaemic myocardium<sup>23,24</sup> and physiological benefit without the formation of stable grafts containing the transplanted cells<sup>25–27</sup>. It has been suggested that engraftment of multipotent cells (which can form many cell types) such as mesenchymal stem cells may carry risks of unwanted cell types forming *in vivo*. One report described bone and cartilage formation in the heart after transplantation<sup>28</sup>, although this could have been calcification resulting from injection of foreign tissue.

When it became apparent that functional improvement is not necessarily correlated with cell engraftment, attention turned to using animal models to investigate the mechanisms involved in the improvement (see ref. 29 for a review). Many cell types, derived from tissues as varied as cord blood, adipose tissue and peripheral blood, behave similarly to bone-marrow cells after being injected directly into the heart or travelling to the ischaemic site after intravenous injection, and therefore any reported improvements in cardiac function are likely to be mediated predominantly by paracrine effects<sup>30</sup>. For example, transplantation of multipotent adult progenitor cells, a cultured bone-marrow-cell-derived population reported to differentiate into derivatives of all germ layers (the three cell layers that give rise to all cells of the adult body), improved cardiac function through the release of inflammatory and vascular growth factors, again in the absence of permanently engrafted cells<sup>31</sup>. Thymosin- $\beta$ 4 (a small actin-binding protein that activates integrin-linked kinase and promotes cardiac cell migration and survival) has been reported to affect the migration of, and vascularization by, an endogenous epithelial progenitor-cell population recently identified in human epicardium<sup>32</sup>. These epithelial cells were as effective in the short term as any others in restoring post-infarct function, reducing dilatation of the heart chambers and increasing ejection fraction in immunodeficient mice receiving the cells<sup>33</sup>. In this context, it is of interest that spontaneous cardiac regeneration in zebrafish after removal of the apex (the lowest part of the heart) is initiated by reprogramming of the epicardium to change its fate and replace all of the missing apex as though it had never been removed<sup>34</sup>.

Taken together, these data indicate that infarcted hearts benefit from transplantation of numerous cell types that are not retained or do not form cardiomyocytes. No adverse events in patients have been linked with transplantation itself, so the procedures seem safe. Therefore, even if the benefits are short term, there may nevertheless be value to patients. Understanding the underlying mechanisms may help in refining factors such as the cell number and cell type to optimize these effects. By contrast, transplantation of cells that do not form grafts may no longer be necessary if paracrine effects can be reproduced by administering the secreted active components.

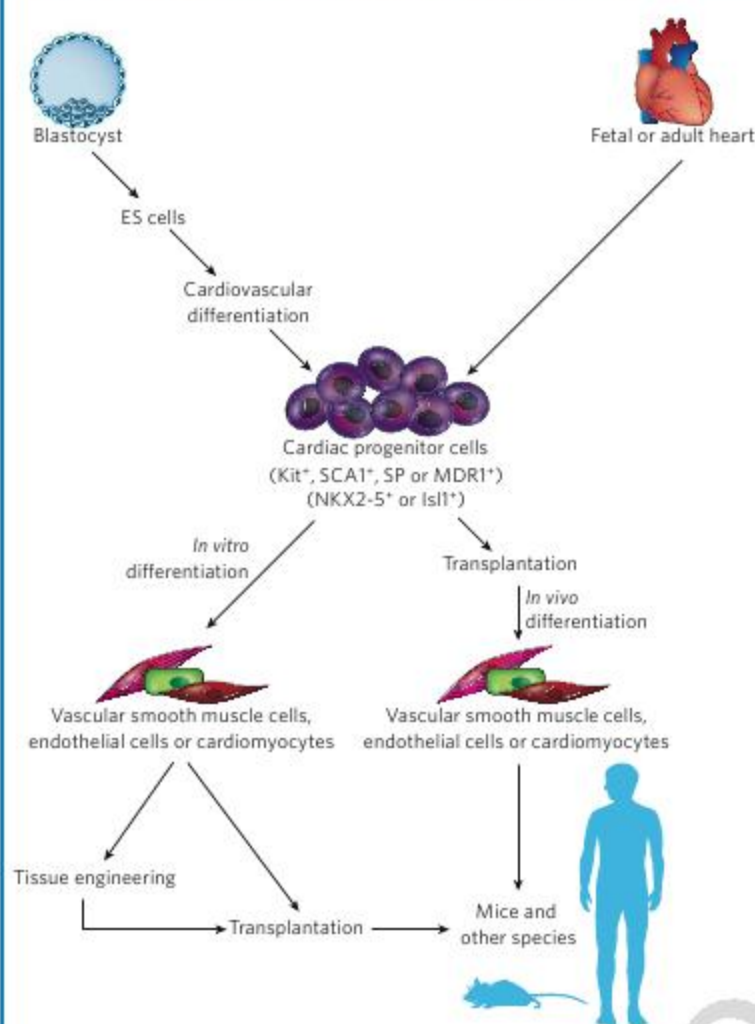
In all studies, the choice of controls for comparison with the test sample is crucial. Whether scientists use no injection, the fluid in which cells are normally suspended alone or control-cell populations that are not relevant to the mechanism under study is likely to affect the outcome and conclusions of any study considerably (Table 1).

### Myocytes and cardiac repair

The great challenge of using cardiomyocytes to replace functional myocardium once it has been destroyed remains. This is especially important for restoring long-term function to prevent or reverse heart failure, which is the main aim of cardiac-cell therapy. The newly formed muscle will provide passive mechanical support but, more importantly, it will



## Box 2 | Cardiac progenitor cells in the fetal and adult heart



The discovery of dividing cells that resemble immature cardiomyocytes in the zone bordering an infarct border zone suggested that the adult heart contains stem cells, although the origin of these cells was unclear<sup>88</sup>. Antonio Beltrami and colleagues<sup>89</sup> subsequently isolated a lineage-negative ( $\text{Lin}^-$ , a cell population that lacks the immunophenotypic antigens that define mature peripheral blood-cell lineages)  $\text{Kit}^+$  cell population from the heart of adult mice that was clonogenic (they are clones of one another), self-renewing and capable of differentiating into cardiomyocytes, vascular smooth muscle cells and endothelial cells both *in vitro* and *in vivo*<sup>89</sup> (see figure). Later, two other groups reported the isolation of adult cardiac stem cells from mice on the basis of expression of stem-cell antigen 1 (SCA1) or exclusion of Hoechst dye (because of the presence of the ATP-binding cassette transporter ABCG2)<sup>90,91</sup>. These dye-excluding cells are also referred to as side population (SP) cells because of their profiles after sorting by flow cytometry. These three populations of cells ( $\text{Kit}^+$  cells,  $\text{SCA1}^+$  cells and SP

cells) from the adult heart differ phenotypically and in their expression of cell-surface markers<sup>92-94</sup>. Their exact relationship to embryonic cardiac progenitor cells that have been found in the fetal heart<sup>6,7,95</sup> is unknown (see ref. 3 for a review).

To determine whether such endogenous cardiac progenitor cells are present in the human heart, scientists isolated a heterogeneous population of cells from biopsied human atrial and ventricular samples by using a mild enzymatic digestion. From the tissue pieces, clonal, multicellular clusters of cells formed in suspension culture, and these are referred to as cardiospheres<sup>96</sup>. Cardiospheres consist of proliferating  $\text{Kit}^+$  stem cells in the core and cells expressing cardiac (myosin heavy chain, troponin I and atrial natriuretic peptide) and endothelial-cell (KDR; also known as FLK1) and CD31) markers on the periphery. In addition, cardiospheres express stem-cell markers CD34 and SCA1 (mouse). Cardiospheres can be passaged and frozen, but the cardiomyocytes derived from these will beat only if they are co-cultured with neonatal rat cardiomyocytes. More recently, cardiosphere-derived cells with improved isolation and expansion efficiencies have been generated from biopsied endomyocardial samples from the right ventricle of adult patients; however, these cells were still isolated as a mixed population without cell sorting or antigenic selection<sup>97</sup>. By contrast, several studies have identified an endogenous  $\text{Kit}^+$  cell population that could be selected from cardiac tissue by using a  $\text{Kit}$ -specific antibody<sup>98</sup>. These cells have been termed human cardiac stem cells. In contrast to human cardiosphere-derived cells, human cardiac stem cells express multidrug resistance protein 1 (MDR1; also known as ABCB1)<sup>99,100</sup>. These studies describe another cardiac progenitor-cell population characterized by the presence of the marker MDR1. It is unclear whether these different stem-cell populations are derived from the heart or are associated with it through the circulation. Apart from MDR1, the main difference between human cardiosphere-derived cells and human cardiac stem cells is that cardiac stem cells express CD45 and CD133 (also known as PROM1). CD34, CD45 and CD133 are all expressed by bone-marrow-derived or endothelial progenitor cells, which might indeed populate the heart by way of the circulation.

The various cell types that can be isolated from fetal and adult hearts by flow-cytometric separation of cells exposed to antibodies specific for the cell-surface proteins SCA1,  $\text{Kit}$  and MDR1 are shown in the figure. Cardiac progenitor cells expressing the cardiac transcription factors NKX2-5 and  $\text{Isl1}$  may be included within these populations (because they are markers of cardiac progenitor cells). Cardiac progenitor cells can also be isolated from embryonic stem cells (human or mouse) by using flow cytometry, if they have been genetically marked at the NKX2-5 and  $\text{Isl1}$  loci.

To repair the heart, these cardiac progenitor cells, as isolated by any of the procedures described, could be transplanted directly into a mouse heart, leading to further differentiation of these cells *in vivo* to become vascular smooth muscle cells, endothelial cells or cardiomyocytes (or a combination of all three cell types). Alternatively, cardiac progenitor cells could be subjected to specific *in vitro* differentiation protocols for generating these cell types, which could then be transplanted directly into the heart or transplanted after tissue engineering into mice and humans (Box 1).

need to couple and contract in synchrony with the host myocardium to be effective and safe.

The only myocytes or myocyte progenitor cells that have been investigated clinically so far are freshly isolated skeletal myoblasts. These had been tested previously in rodent myocardial-infarction models, in much the same way as the non-cardiomyocytes described above. The clinical use of skeletal myoblasts, however, has been limited even though their administration improved the ejection fraction, because the myocytes that formed were unable to couple functionally with the host myocardium and may therefore cause arrhythmia (abnormalities in heart rhythm)<sup>35</sup>. Skeletal myoblasts and myogenic precursor/progenitor cells normally mediate development and regeneration of skeletal muscle<sup>36</sup>. A recent study showed, however, that skeletal myoblasts could be engineered to express the gap-junction protein connexin 43 (ref. 37). This modification improved electrical coupling between the

infarct region and the surrounding myocardium. Similar approaches have been suggested to drive cardiomyocytes towards electrophysiological maturation<sup>38</sup>, although this could also be a disadvantage for intramyocardial transplantation because (immature) fetal myocytes survive better after injection in the heart than adult cardiomyocytes<sup>39</sup>. It does, however, seem to be important that grafted cells are muscle cells and not myofibroblasts, which can cause ectopic electrical activity, increasing the risk of arrhythmia<sup>40</sup>.

### Human ES cells as a source of cardiomyocytes

Of all stem-cell sources, human ES cells are among the more promising for realizing cardiac regeneration through cardiomyocyte replacement because of their capacity to undergo directed differentiation into genuine cardiomyocytes and supportive cardiac cells<sup>41-43</sup> *in vitro* fairly efficiently. These *in vitro*-derived cells have been characterized extensively<sup>44-49</sup>.



**Table 1 | Cell-transplantation studies in models of experimental myocardial infarction**

Cell type*	Number ( $\times 10^6$ )	Recipient species	Time after infarction	Measurement of function	Result	Longest follow-up	Control	Reference
BMCs	0.03–0.5	Mouse	3–5 hours	Pressure–volume loop	↓ LVEDP, ↑ LVSP, ↑ dP/dt	1–2 weeks	No injection or Kit <sup>+</sup> BMCs (control groups pooled)	9
	1	Macaque	Directly	Echocardiography	↑ FS	2 weeks	Saline	24
Mesenchymal stem cells	2	Rat	7 days	Echocardiography	↑ FS, ↓ EDV	4 weeks	Saline	27
	50	Pig	Directly	MRI, nuclear imaging	↑ EF, ↓ ESV	4 weeks	No injection	26
Multipotent adult progenitor cells	0.5	Mouse	Directly	Echocardiography	↑ EF, ↑ FS	8 weeks	BMCs, fibroblasts or medium	31
	50	Pig	Directly	MRI, nuclear imaging	↑ EF, ↑ bioenergetics	4 weeks	Saline	77
EPDCs (human)	0.4	Mouse	Directly	MRI	↑ EF, ↓ EDV	6 weeks	Medium	33
Skeletal myoblasts (50% pure)	10	Rat	7 days	Treadmill, <i>ex vivo</i> pressure–volume loop	↓ EDV, ↑ LVSP, ↑ exercise tolerance	6 weeks	Saline	78
Skeletal myoblasts	150–750	Sheep	14 days	Echocardiography	↑ EF, ↓ EDV	16 weeks	Medium	79
ES cells committed to cardiac lineage (mouse)	50	Sheep	14 days	Echocardiography	↑ EF (relative change over time)	4 weeks	Medium	80
ES-cell-derived cardiomyocytes	0.03–0.1	Mouse	Directly	Pressure–volume loop	↑ EF, ↓ EDV	3–4 weeks	BMCs, fibroblasts or medium	81
ES-cell-derived cardiomyocytes (human, 25% pure)	1	Mouse	Directly	MRI	↑ EF (4 weeks), ↔ EF (12 weeks)	12 weeks	ES-cell-derived non-cardiomyocytes (human)	22
ES-cell-derived cardiomyocytes (human, 71–95% pure)	10	Rat	4 days	Echocardiography, MRI	↑ FS, ↓ EDV, ↑ wall thickening	4 weeks	ES-cell-derived non-cardiomyocytes (human) or saline	54
ES-cell-derived cardiomyocytes (human, <10% pure)	1.5	Rat	7–10 days	Echocardiography	↑ FS, ↓ EDV	4 weeks	ES-cell-derived non-cardiomyocytes (human) or saline	82

BMCs, bone-marrow cells; dP/dt, change in systolic pressure over time; EDV, end diastolic volume; EF, ejection fraction; ESV, end systolic volume; EPDCs, epicardium-derived stem cells; FS, fractional shortening (shortening of the ventricular diameter during systole; it indicates the degree of contraction); LVEDP, left ventricular end diastolic pressure; LVSP, left ventricular systolic pressure; medium, cell-culture medium; MRI, magnetic resonance imaging; nuclear imaging modality, <sup>31</sup>P magnetic resonance spectroscopy; wall thickening, thickening of the ventricular wall. \*Unless otherwise specified, transplanted cells are from the same species. If the purity is not listed, then it is almost pure.

The first reports of transplantation of human ES-cell-derived cardiomyocytes into pigs and guinea pigs demonstrated their potential to function as biological pacemakers in electrophysiologically silenced or AV-blocked hearts, in which conductance between the atrium and ventricle has been blocked<sup>50,51</sup>. These experiments required relatively few human ES-cell-derived cardiomyocytes (only hundreds) to exert their effect. Recent advances in generating human ES-cell-derived cardiomyocytes in higher numbers (millions) mean that it is now possible to investigate their regenerative potential. After transplantation into the healthy myocardium of immunodeficient rats or mice, human ES-cell-derived cardiomyocytes survived<sup>22,52,53</sup> and matured<sup>22,53</sup> for at least 12 weeks<sup>22</sup>. In all studies so far, mixtures of cardiomyocytes and other differentiated human ES-cell derivatives have been injected into the heart. Where this has been examined in detail, preferential survival of cardiomyocytes over non-cardiomyocytes is usually observed, with non-cardiac cells lost over time without being detected in any other organs. Although grafted cells formed a syncytium, or connected heart tissue, with each other, they were usually separated from the rodent myocardium by a layer of fibrotic tissue. Reducing the secretion of extracellular-matrix proteins, which form the basis of fibrotic tissue, seems to be important to facilitate proper electrical coupling, because the presence of fibrotic patches is a major risk factor for arrhythmia. In this respect, it is notable that the transplanted cells themselves contribute to their isolation by secreting particular extracellular-matrix components (Fig. 2).

When transplanted into the infarcted heart, human ES-cell-derived cardiomyocytes are the only cells that form grafts visible to the naked eye (Fig. 2), although in one study this depended on the addition of a pro-survival cocktail<sup>54</sup>, a mixture of proteins that prevented the cardiomyocytes from dying after transplantation. Cardiac function was improved at 4 weeks after transplantation in mice receiving human ES-cell-derived cardiomyocytes compared with those receiving non-cardiomyocyte derivatives<sup>22,54,55</sup>. However, the rodents receiving human ES-cell-derived cells without cardiomyocytes also showed some improvement compared with animals that received only the fluid in

which cells are normally suspended. Hence, cardiomyocyte-specific benefits were detected beyond those from non-cardiomyocyte cell types, but as a cautionary note, if the vehicle alone had been taken as reference, the relative functional improvements would have seemed much higher.

Only one study so far has extended the analysis of heart function after human ES-cell transplantation up to 12 weeks in mice (Table 1). At this time point, the benefit of receiving cardiomyocytes derived from human ES cells (in comparison to non-cardiomyocytes) was no longer present<sup>22</sup>, although it was still higher than for the vehicle control. Functional enhancement by human ES-cell-derived cardiomyocytes with the current strategies may thus be limited to mid-term at most, equivalent to perhaps a few months in humans, and there is no evidence that the underlying mechanism depends on the contractile properties of the transplanted cells. The paracrine mechanism that has been proposed for adult stem cells may also be applicable to human ES-cell-derived cardiomyocytes. Parameters that warrant further investigation include ways to limit extracellular-matrix formation around the graft and ways to allow later injection of cells, for example after the initial inflammatory phase when the environment may be hostile to the donor cells. The present discrepancy between mid-term and long-term results and the dependence of outcome on the particular control argues for careful re-evaluation of reported improvements by any cell type<sup>56</sup>.

An important parameter in evaluating the function of human cardiomyocytes in rodents is the difference in their intrinsic beating frequency. Human cardiomyocytes beating 60–100 times per minute are transplanted into a rodent heart beating 300–600 times per minute. Unsuccessful coupling, as demonstrated by intracellular calcium imaging<sup>57,58</sup>, could contribute to the failure to improve long-term cardiac function. By contrast, prolonged periods of tachycardia (rapid beating of the heart) from coupling of human ES-cell-derived cardiomyocytes to host cardiac cells may induce cardiac dysfunction. Experiments in larger animals such as pigs or primates with lower heart rates are needed to resolve the issue. These would preferably be ES cells of the same species or iPS-cell derivatives, which could be autologous (Box 1).



Studies in large animals simultaneously will be useful for assessing the risk of tumour formation by human ES cell and iPS cells because this may be species dependent<sup>59</sup>.

Although clinical studies using cells from different sources for transplantation are actively being pursued, it is clear that many basic questions remain unanswered. Apart from understanding the mechanistic basis underlying potentially beneficial effects, the best cell type for each application (for example, for acute myocardial infarction versus chronic heart failure), the timing of cell therapy and the delivery methods in most cases still need to be determined. Tissue engineering<sup>60–62</sup>, combining tissue cells with artificial or natural scaffolds, or intramyocardial injection of combinations of cell types (such as cardiomyocytes and angiogenic cells) seem more appropriate than intracoronary injection of single-cell suspensions because the latter need to align correctly and may cause micro-infarctions<sup>63</sup>, but much more research is needed before clinical strategies can be properly defined.

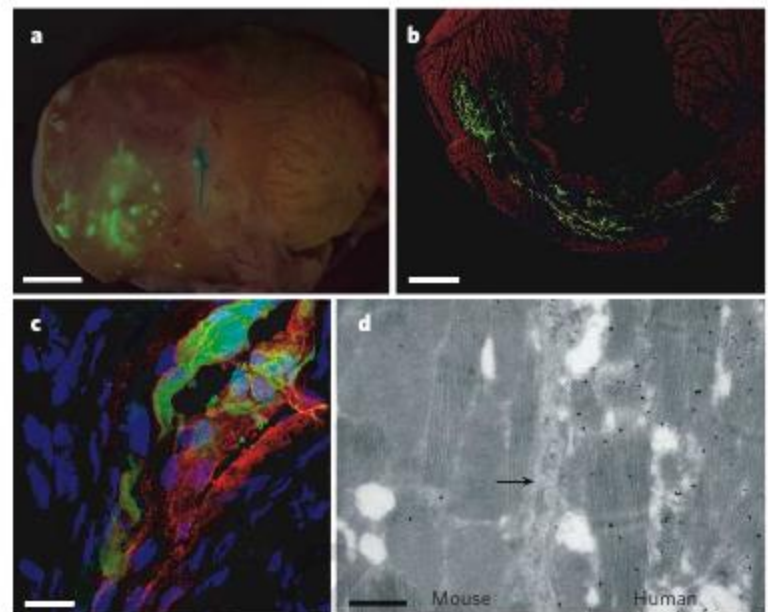
### Transplanting differentiated non-cardiomyocytes

Integration of grafts into host tissue requires the formation of blood vessels to supply oxygen and nutrients. Shulamit Levenberg and colleagues<sup>43</sup> induced human ES cells to differentiate into endothelial cells, isolated them by using antibodies specific for platelet endothelial cell-adhesion molecule 1 and demonstrated their ability to form vascular-like structures after subcutaneous transplantation in a biodegradable scaffold in mice. Mouse blood cells were subsequently detected in some of the human blood vessels, suggesting that anastomosis, or the formation of connections, occurred between mouse and human microvessels. This integration of human blood vessels into the host circulation was also observed in other studies using human ES-cell-derived endothelial cells and smooth muscle cells<sup>41,64</sup>, even up to 5 months after transplantation<sup>64</sup>. In a model of hindlimb ischaemia in mice, an increased number of capillaries and improved perfusion (blood-flow) rates were also observed 2 and 4 weeks after transplantation with human ES-cell-derived endothelial cells<sup>65,66</sup> or endothelial cells and smooth muscle cells together<sup>67</sup>. Transplantation of haemangioblasts derived from human ES cells resulted in haematopoietic cells, as well as endothelial cells, and contributed to repair of damaged vasculature<sup>68</sup>. Human ES cells have also been shown to form cells that are similar to haematopoietic stem cells, with the capacity to form cells of the haematopoietic lineage<sup>69,70</sup>.

Vascular cells are important in myocardial repair for two reasons. First, neovascularization protects cardiomyocytes surrounding the infarct, helping to retain cardiac function. Mice with a mutated transforming growth factor- $\beta$  receptor gene, endoglin, leading to mononuclear-cell defects, suffered from impaired vascularization and cardiac function after myocardial infarction. The defects could be rescued by intravenous injection of healthy donor mononuclear cells but not mononuclear cells from patients with the same gene mutation<sup>71</sup>. These data indicate that patients with heart disease and related vascular impairments, whose mononuclear cells and endothelial progenitor cells often have impaired function, may benefit from transplantation of healthy endothelial cells, either from matched, heterologous human ES cells or from autologous endothelial progenitor cells. Second, transplanted cardiomyocytes themselves need an additional blood supply, particularly if fibrosis tends to isolate grafts from the host vasculature. Co-transplanted vascular cells may thus help connect the grafts to the existing vascular network. Alternatively, bi-potent or tri-potent cardiac progenitor cells (Box 2) from human ES cells, or fetal or adult heart, may be able to form cardiomyocytes, smooth muscle cells and endothelial cells *in situ*.

### Transplanting differentiated human ES cells into other tissues

The central nervous system is another tissue site that has been targeted for development of cell-transplantation therapies. One reason for this is that neural differentiation of human ES cells has been relatively straightforward. Many human ES cell lines spontaneously differentiate into neurons when grown in the absence of supporting feeder fibroblasts normally present to support growth and inhibit



**Figure 2 | A graft derived from human ES cells and extracellular-matrix production in a mouse heart.** **a**, A mouse heart 12 weeks after the induction of a myocardial infarction and injection of hES3-GFP-derived cardiomyocytes (green). Scale bar, 250  $\mu$ m. **b**, A transverse section of infarcted left ventricle 3 weeks after ligation of the coronary artery and transplantation of human ES3-GFP-derived cardiomyocytes (green). Cardiomyocytes are characterized by staining with an antibody specific for tropomyosin. Scale bar, 300  $\mu$ m. **c**, A section of infarcted left ventricle 12 weeks after injection of human ES3-GFP-derived cardiomyocytes (green), showing the production of human collagen type IV (red). Cell nuclei are also shown (blue). Scale bar, 20  $\mu$ m. **d**, An electron-microscopic image of human ES3-GFP-derived cardiomyocytes 12 weeks after transplantation in a mouse heart. Immunogold labelling of GFP (black dots) was used to distinguish human and mouse cardiac cells. Extracellular-matrix (see arrow) proteins surround the graft, causing electrical isolation and failure of the transplanted cells to connect to the host (mouse) heart tissue. Scale bar, 500 nm. Red indicates tropomyosin; blue indicates DNA; and yellow indicates GFP and tropomyosin in the same cell. (Panel d courtesy of A. Verkleij and E. van Donselaar, Utrecht University, The Netherlands.)

differentiation. Furthermore, in common with cardiomyocytes, mature neurons are terminally differentiated cells that cannot proliferate and potentially could replace dead or malfunctioning cells in the central nervous system. The loss of single or multiple neural cell types causes a broad range of neurological disorders such as spinal-cord lesions, Parkinson's disease (in which there is a loss of neurons that produce the neurotransmitter dopamine), amyotrophic lateral sclerosis (Lou Gehrig's disease, also known as motor neuron disease), Huntington's disease (a genetic disorder that results in neuronal cell death) and age-related macular degeneration (a disease causing blindness). These diseases are thought to be good candidates for cell-transplantation therapy. Just as for cell therapy to treat cardiovascular diseases, the long-term survival, functional integration and physiological compatibility of engrafted cells are important parameters for successful treatment. Some studies have shown long-term (greater than 3 months) survival of transplanted human ES-cell-derived neurons or neuronal progenitor cells, in particular in rodent models of Parkinson's disease, although cell numbers were low and functional recovery was only partial. The treatment of neurological disorders with human ES-cell-derived cells and various other sources, including neural stem cells or progenitor cells from fetal brain, are reviewed elsewhere<sup>72</sup>.

Functional integration of transplanted cells with host tissue may not necessarily be required for treatment of some diseases, such as liver diseases and diabetes, the latter of which is characterized by the loss of insulin-producing pancreatic  $\beta$ -cells (type 1 diabetes) or by reduced sensitivity to insulin (type 2 diabetes). Efficient blood supply and, in the case of pancreatic cells, release of insulin in response to glucose, may be sufficient for treatment of the disease. Transplantation of human ES-cell-derived



pancreatic cells into the subcapsular renal space in streptozotocin-treated diabetic mice decreased glucose levels in the blood and increased body weight. On removal of the graft, these effects were reversed<sup>73</sup>. This study supports the idea that transplanted pancreatic cells do not need to integrate into pancreatic tissue to be functional.

### Towards the future

In the years immediately after the first derivation of human ES cells, research focused on optimizing differentiation protocols for the cell type of choice, but more recently the focus has shifted towards transplantation therapies. Advances in the generation of iPS cells will probably prompt a further increase in the number of transplantation studies. However, for successful translation from experimental studies to clinical trials, many problems have to be solved. In particular, in conditions where multiple cell types have to be replaced and functional integration is necessary, we can learn valuable lessons from transplantation studies in the heart. It has become clear that restoration of heart function by replacing myocardium requires long-term integration and establishment of functional connections with the host tissue, in all cases using multiple controls that include non-functional cells, cell-free solution in which cells are normally suspended, immunosuppression alone and mock injections.

Animal experiments should include strategies to prove that transplanted cells actually cause the effect observed. These could be through inclusion of constructs that contain conditionally expressed suicide genes. These could be activated after their function has been restored with the expectation that the functional benefit would be reversed. Alternatively, in specific situations, grafted tissue could be removed after recovery, as was done in the transplantation of pancreatic cells<sup>73</sup>. Most transplantation studies have used single-cell suspensions or small groups of cells. This is generally accompanied by extensive cell death and inadequate integration. Particularly in a hostile immunoreactive, ischaemic or necrotic environment, survival of the grafted tissue is a major challenge. Instead of direct injection of cells into the site of injury or disease, an alternative delivery method could be to seed cells *ex vivo* on a biodegradable polymeric scaffold, followed by *in vivo* engraftment. Three-dimensional multicellular culture systems may also have the advantage of supporting cell proliferation, maturation and organization of cells more effectively. The disease or injury in question will probably determine whether injection of cells or grafting tissue-engineered constructs is the preferred choice<sup>67</sup>.

In addition to improving delivery methods, it will be important that, despite recent advances, further progress is made in refining effective differentiation protocols that yield highly purified cell populations for transplantation. This could best be achieved by cell sorting using cell-type-specific membrane markers, although genetic methods using cell-type-specific promoter constructs driving, for example, GFP<sup>74–76</sup> or selected survival by antibiotic resistance could be alternatives for some purposes. As well as increasing the density of the cell type of interest at the site of transplantation, the use of highly purified populations will also minimize the chance of co-transplanting contaminating undifferentiated stem cells and hence reduce the risk of developing a teratoma. Furthermore, the optimal differentiation state of cells for successful transplantation remains to be determined. To this end, lineage-specific and cell-type-specific promoters could be used to select progenitor cells or more differentiated cells for comparison of effectiveness and safety in animal models. Approaches to identifying embryonic progenitor cells in mice<sup>67</sup> may be applied to ES cells and may help to uncover new ways to isolate human ES cells and human iPS cells cleanly. Additional work in mouse ES cells will clearly be required before human ES-cell-based therapy can be launched in the clinic.

In general, mature cells have a lower proliferation rate and a lower survival rate after transplantation than progenitor cells. In addition, the microenvironment of transplanted cells often provides cues for further differentiation into mature cells. These issues need to be elaborated in future transplantation studies, which will lead to fine-tuning of transplantation therapy and eventually to repair of diseased or damaged organs. ■

- Thomson, J. A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Draper, J. S. et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature Biotechnol.* **22**, 53–54 (2004).
- Wu, S., Chien, K. & Mummery, C. Origin and biology of multipotent cardiovascular progenitor cells. Reverse translational medicine towards models of human heart disease. *Cell* **132**, 537–543 (2008).  
This review describes the origin and fate of cardiac progenitor cells in embryonic and adult hearts, as well as those derived from ES cells.
- Massberg, S. et al. Immunosurveillance by hematopoietic progenitor cells trafficking through blood, lymph, and peripheral tissues. *Cell* **131**, 994–1008 (2007).
- Pouly, J. et al. Cardiac stem cells in the real world. *J. Thorac. Cardiovasc. Surg.* **135**, 673–678 (2007).
- Wu, S. M. et al. Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127**, 1137–1150 (2006).  
This paper describes the molecular identity of Kit<sup>+</sup>NKX2-5<sup>+</sup> progenitor cells in the heart and traces their fate during development in mice.
- Moretti, A. et al. Multipotent embryonic *Isl1*<sup>+</sup> progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151–1165 (2006).  
This paper describes genetic fate-mapping studies and shows that expression of *Isl1*, *Nkx2-5* and *Kdr* defines multipotent cardiovascular progenitor cells, which can give rise to endothelial cells, cardiomyocytes and smooth muscle cells.
- Brazelton, T. R., Rossi, F. M., Keshet, G. I. & Blau, H. M. From marrow to brain: expression of neuronal phenotypes in adult mice. *Science* **290**, 1775–1779 (2000).
- Orlic, D. et al. Bone marrow cells regenerate infarcted myocardium. *Nature* **410**, 701–705 (2001).
- Guan, K. & Hasenfuss, G. Do stem cells in the heart truly differentiate into cardiomyocytes? *J. Mol. Cell. Cardiol.* **43**, 377–387 (2007).
- Assmus, B. et al. Transcatheter transplantation of progenitor cells after myocardial infarction. *N. Engl. J. Med.* **355**, 1222–1232 (2006).
- Schachinger, V. et al. Intracoronary bone marrow-derived progenitor cells in acute myocardial infarction. *N. Engl. J. Med.* **355**, 1210–1221 (2006).
- Lunde, K. et al. Intracoronary injection of mononuclear bone marrow cells in acute myocardial infarction. *N. Engl. J. Med.* **355**, 1199–1209 (2006).
- Janssens, S. et al. Autologous bone marrow-derived stem-cell transfer in patients with ST-segment elevation myocardial infarction: double-blind, randomised controlled trial. *Lancet* **367**, 113–121 (2006).
- Cintron, G., Johnson, G., Francis, G., Cobb, F. & Cohn, J. N. Prognostic significance of serial changes in left ventricular ejection fraction in patients with congestive heart failure. *Circulation* **87**, VI17–VI23 (1993).
- Adler, E. D. & Maddox, T. M. Cell therapy for cardiac disease: where do we go from here? *Nature Clin. Pract. Cardiovasc. Med.* **4**, 2–3 (2007).
- Lunde, K. et al. Exercise capacity and quality of life after intracoronary injection of autologous mononuclear bone marrow cells in acute myocardial infarction: results from the Autologous Stem cell Transplantation in Acute Myocardial Infarction (ASTAMI) randomized controlled trial. *Am. Heart J.* **154**, 710–718 (2007).
- Schachinger, V., Tonn, T., Dimmeler, S. & Zeiher, A. M. Bone-marrow-derived progenitor cell therapy in need of proof of concept: design of the REPAIR-AMI trial. *Nature Clin. Pract. Cardiovasc. Med.* **3** (Suppl. 1), S23–S28 (2006).
- Murry, C. E. et al. Haematopoietic stem cells do not transdifferentiate into cardiac myocytes in myocardial infarcts. *Nature* **428**, 664–668 (2004).
- Balsam, L. B. et al. Haematopoietic stem cells adopt mature haematopoietic fates in ischaemic myocardium. *Nature* **428**, 668–673 (2004).  
References 19 and 20 were the first studies to show that haematopoietic stem cells do not transdifferentiate into cardiomyocytes when transplanted into a mouse heart.
- Nygren, J. M. et al. Bone marrow-derived hematopoietic cells generate cardiomyocytes at a low frequency through cell fusion, but not transdifferentiation. *Nature Med.* **10**, 494–501 (2004).
- van Laake, L. W. et al. Human embryonic stem cell-derived cardiomyocytes survive and mature in the mouse heart and transiently improve function after myocardial infarction. *Stem Cell Res.* **1**, 9–24 (2007).  
This paper showed an improvement in cardiac function at 4 weeks after transplantation of human ES-cell-derived cardiomyocytes into mice that had undergone a myocardial infarction; however, this effect was not sustained at 12 weeks compared with mice receiving human ES-cell-derived non-cardiomyocytes.
- Fazel, S. et al. Cardioprotective c-kit<sup>+</sup> cells are from the bone marrow and regulate the myocardial balance of angiogenic cytokines. *J. Clin. Invest.* **116**, 1865–1877 (2006).
- Yoshioka, T. et al. Repair of infarcted myocardium mediated by transplanted bone marrow-derived CD34<sup>+</sup> stem cells in a nonhuman primate model. *Stem Cells* **23**, 355–364 (2005).
- Limbourg, F. P. et al. Haematopoietic stem cells improve cardiac function after infarction without permanent cardiac engraftment. *Eur. J. Heart Fail.* **7**, 722–729 (2005).
- Feygin, J., Mansoor, A., Eckman, P., Swigen, C. & Zhang, J. Functional and bioenergetic modulations in the infarct border zone following autologous mesenchymal stem cell transplantation. *Am. J. Physiol. Heart Circ. Physiol.* **293**, H1772–H1780 (2007).
- Amsalem, Y. et al. Iron-oxide labeling and outcome of transplanted mesenchymal stem cells in the infarcted myocardium. *Circulation* **116**, 138–145 (2007).
- Breitbach, M. et al. Potential risks of bone marrow cell transplantation into infarcted hearts. *Blood* **110**, 1362–1369 (2007).
- Rubart, M. & Field, L. J. Cardiac regeneration: repopulating the heart. *Annu. Rev. Physiol.* **68**, 29–49 (2006).
- Gnecchi, M. et al. Paracrine action accounts for marked protection of ischemic heart by Akt-modified mesenchymal stem cells. *Nature Med.* **11**, 367–368 (2005).
- Pelacho, B. et al. Multipotent adult progenitor cell transplantation increases vascularity and improves left ventricular function after myocardial infarction. *J. Tissue Eng. Regen. Med.* **1**, 51–59 (2007).
- Smart, N. et al. Thymosin  $\beta$ 4 induces adult epicardial progenitor mobilization and neovascularization. *Nature* **445**, 177–182 (2007).



33. Winter, E. M. *et al.* Preservation of left ventricular function and attenuation of remodeling after transplantation of human epicardium-derived cells into the infarcted mouse heart. *Circulation* **116**, 917–927 (2007).
34. Lepilina, A. *et al.* A dynamic epicardial injury response supports progenitor cell activity during zebrafish heart regeneration. *Cell* **127**, 607–619 (2006).
35. Leobon, B. *et al.* Myoblasts transplanted into rat infarcted myocardium are functionally isolated from their host. *Proc. Natl Acad. Sci. USA* **100**, 7808–7811 (2003).
36. Dellavalle, A. *et al.* Pericytes of human skeletal muscle are myogenic precursors distinct from satellite cells. *Nature Cell Biol.* **9**, 255–267 (2007).
37. Roell, W. *et al.* Engraftment of connexin 43-expressing cells prevents post-infarct arrhythmia. *Nature* **450**, 819–824 (2007).
38. Liu, J., Fu, J. D., Siu, C. W. & Li, R. A. Functional sarcoplasmic reticulum for calcium-handling of human embryonic stem cell-derived cardiomyocytes: Insights for driven maturation. *Stem Cells* **25**, 3038–3044 (2007).
39. Reinecke, H., Zhang, M., Bartosek, T. & Murry, C. E. Survival, integration, and differentiation of cardiomyocyte grafts: a study in normal and injured rat hearts. *Circulation* **100**, 193–202 (1999).
40. Miragoli, M., Salvarani, N. & Rohr, S. Myofibroblasts induce ectopic activity in cardiac tissue. *Circ. Res.* **101**, 755–758 (2007).
41. Ferreira, L. S. *et al.* Vascular progenitor cells isolated from human embryonic stem cells give rise to endothelial and smooth muscle like cells and form vascular networks in vivo. *Circ. Res.* **101**, 286–294 (2007).
42. Huang, H. *et al.* Differentiation of human embryonic stem cells into smooth muscle cells in adherent monolayer culture. *Biochem. Biophys. Res. Commun.* **351**, 321–327 (2006).
43. Levenberg, S., Golub, J. S., Amit, M., Itskovitz-Eldor, J. & Langer, R. Endothelial cells derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **99**, 4391–4396 (2002).
44. Kehat, I. *et al.* Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J. Clin. Invest.* **108**, 407–414 (2001).  
**This was the first paper to describe the formation of cardiomyocytes from human ES cells.**
45. Mummery, C. *et al.* Differentiation of human embryonic stem cells to cardiomyocytes: role of coculture with visceral endoderm-like cells. *Circulation* **107**, 2733–2740 (2003).
46. He, J. Q., Ma, Y., Lee, Y., Thomson, J. A. & Kamp, T. J. Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization. *Circ. Res.* **93**, 32–39 (2003).
47. Xu, C., Police, S., Rao, N. & Carpenter, M. K. Characterization and enrichment of cardiomyocytes derived from human embryonic stem cells. *Circ. Res.* **91**, 501–508 (2002).
48. Beqqali, A., Kloots, J., Ward-van Oostwaard, D., Mummery, C. & Passier, R. Genome-wide transcriptional profiling of human embryonic stem cells differentiating to cardiomyocytes. *Stem Cells* **24**, 1956–1967 (2006).
49. Passier, R. *et al.* Increased cardiomyocyte differentiation from human embryonic stem cells in serum-free cultures. *Stem Cells* **23**, 772–780 (2005).
50. Kehat, I. *et al.* Electromechanical integration of cardiomyocytes derived from human embryonic stem cells. *Nature Biotechnol.* **22**, 1282–1289 (2004).
51. Xue, T. *et al.* Functional integration of electrically active cardiac derivatives from genetically engineered human embryonic stem cells with quiescent recipient ventricular cardiomyocytes: insights into the development of cell-based pacemakers. *Circulation* **111**, 11–20 (2005).
52. Laflamme, M. A. *et al.* Formation of human myocardium in the rat heart from human embryonic stem cells. *Am. J. Pathol.* **167**, 663–671 (2005).
53. Dai, W. *et al.* Survival and maturation of human embryonic stem cell-derived cardiomyocytes in rat hearts. *J. Mol. Cell Cardiol.* **43**, 504–516 (2007).
54. Laflamme, M. A. *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnol.* **25**, 1015–1024 (2007).  
**This paper showed that rodents that had undergone a myocardial infarction had improved cardiac function at 4 weeks after transplantation of human ES-cell-derived cardiomyocytes.**
55. Leor, J. *et al.* Human embryonic stem cell transplantation to repair the infarcted myocardium. *Heart* **93**, 1278–1284 (2007).
56. van Laake, L. W. *et al.* Monitoring of cell therapy and assessment of cardiac function using magnetic resonance imaging in a mouse model of myocardial infarction. *Nature Protocols* **2**, 2551–2567 (2007).
57. Rubart, M., Wang, E., Dunn, K. W. & Field, L. J. Two-photon molecular excitation imaging of Ca<sup>2+</sup> transients in Langendorff-perfused mouse hearts. *Am. J. Physiol. Cell Physiol.* **284**, C1654–C1668 (2003).
58. Rubart, M. *et al.* Physiological coupling of donor and host cardiomyocytes after cellular transplantation. *Circ. Res.* **92**, 1217–1224 (2003).
59. Erdo, F. *et al.* Host-dependent tumorigenesis of embryonic stem cell transplantation in experimental stroke. *J. Cereb. Blood Flow Metab.* **23**, 780–785 (2003).
60. Zimmermann, W. H. *et al.* Engineered heart tissue grafts improve systolic and diastolic function in infarcted rat hearts. *Nature Med.* **12**, 452–458 (2006).  
**This paper showed that tissue engineering is an attractive prospect for cardiac repair.**
61. Feinberg, A. W. *et al.* Muscular thin films for building actuators and powering devices. *Science* **317**, 1366–1370 (2007).
62. Furuta, A. *et al.* Pulsatile cardiac tissue grafts using a novel three-dimensional cell sheet manipulation technique functionally integrates with the host heart, in vivo. *Circ. Res.* **98**, 705–712 (2006).
63. Moelker, A. D. *et al.* Intracoronary delivery of umbilical cord blood derived unrestricted somatic stem cells is not suitable to improve LV function after myocardial infarction in swine. *J. Mol. Cell Cardiol.* **42**, 735–745 (2007).
64. Wang, Z. Z. *et al.* Endothelial cells derived from human embryonic stem cells form durable blood vessels in vivo. *Nature Biotechnol.* **25**, 317–318 (2007).
65. Sone, M. *et al.* Pathway for differentiation of human embryonic stem cells to vascular cell components and their potential for vascular regeneration. *Arterioscler. Thromb. Vasc. Biol.* **27**, 2127–2134 (2007).
66. Cho, S. W. *et al.* Improvement of postnatal neovascularization by human embryonic stem cell derived endothelial-like cell transplantation in a mouse model of hindlimb ischemia. *Circulation* **116**, 2409–2419 (2007).
67. Caspi, O. *et al.* Tissue engineering of vascularized cardiac muscle from human embryonic stem cells. *Circ. Res.* **100**, 263–272 (2007).
68. Lu, S. J. *et al.* Generation of functional hemangioblasts from human embryonic stem cells. *Nature Methods* **4**, 501–509 (2007).
69. Tian, X., Woll, P. S., Morris, J. K., Linehan, J. L. & Kaufman, D. S. Hematopoietic engraftment of human embryonic stem cell-derived cells is regulated by recipient innate immunity. *Stem Cells* **24**, 1370–1380 (2006).
70. Narayan, A. D. *et al.* Human embryonic stem cell-derived hematopoietic cells are capable of engrafting primary as well as secondary fetal sheep recipients. *Blood* **107**, 2180–2183 (2006).
71. van Laake, L. W. *et al.* Endoglin has a crucial role in blood cell-mediated vascular repair. *Circulation* **114**, 2288–2297 (2006).
72. Goldman, S. Stem and progenitor cell-based therapy of the human central nervous system. *Nature Biotechnol.* **23**, 862–871 (2005).
73. Shim, J. H. *et al.* Directed differentiation of human embryonic stem cells towards a pancreatic cell fate. *Diabetologia* **50**, 1228–1238 (2007).
74. Duan, Y. *et al.* Differentiation and enrichment of hepatocyte-like cells from human embryonic stem cells in vitro and in vivo. *Stem Cells* **25**, 3058–3068 (2007).
75. Huber, I. *et al.* Identification and selection of cardiomyocytes during human embryonic stem cell differentiation. *FASEB J.* **21**, 2551–2563 (2007).
76. Anderson, D. *et al.* Transgenic enrichment of cardiomyocytes from human embryonic stem cells. *Mol. Ther.* **15**, 2027–2036 (2007).
77. Zeng, L. *et al.* Bioenergetic and functional consequences of bone marrow-derived multipotent progenitor cell transplantation in hearts with postinfarction left ventricular remodeling. *Circulation* **115**, 1866–1875 (2007).
78. Jain, M. *et al.* Cell therapy attenuates deleterious ventricular remodeling and improves cardiac performance after myocardial infarction. *Circulation* **103**, 1920–1927 (2001).
79. Ghostine, S. *et al.* Long-term efficacy of myoblast transplantation on regional structure and function after myocardial infarction. *Circulation* **106**, 1131–1136 (2002).
80. Menard, C. *et al.* Transplantation of cardiac-committed mouse embryonic stem cells to infarcted sheep myocardium: a preclinical study. *Lancet* **366**, 1005–1012 (2005).
81. Kolossov, E. *et al.* Engraftment of engineered ES cell-derived cardiomyocytes but not BM cells restores contractile function to the infarcted myocardium. *J. Exp. Med.* **203**, 2315–2327 (2006).
82. Caspi, O. *et al.* Transplantation of human embryonic stem cell-derived cardiomyocytes improves myocardial performance in infarcted rat hearts. *J. Am. Coll. Cardiol.* **50**, 1884–1893 (2007).
83. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
84. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
85. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
86. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26**, 101–106 (2007).
87. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318**, 1920–1923 (2007).
88. Beltrami, A. P. *et al.* Evidence that human cardiac myocytes divide after myocardial infarction. *N. Engl. J. Med.* **344**, 1750–1757 (2001).
89. Beltrami, A. P. *et al.* Adult cardiac stem cells are multipotent and support myocardial regeneration. *Cell* **114**, 763–776 (2003).
90. Oh, H. *et al.* Cardiac progenitor cells from adult myocardium: homing, differentiation, and fusion after infarction. *Proc. Natl Acad. Sci. USA* **100**, 12313–12318 (2003).
91. Martin, C. M. *et al.* Persistent expression of the ATP-binding cassette transporter, Abcg2, identifies cardiac SP cells in the developing and adult heart. *Dev. Biol.* **265**, 262–275 (2004).
92. Murry, C. E., Reinecke, H. & Pabon, L. M. Regeneration gaps: observations on stem cells and cardiac repair. *J. Am. Coll. Cardiol.* **47**, 1777–1785 (2006).
93. Evans, S. M., Mummery, C. & Doevendans, P. A. Progenitor cells for cardiac repair. *Semin. Cell Dev. Biol.* **18**, 153–160 (2007).
94. Parmacek, M. S. & Epstein, J. A. Pursuing cardiac progenitors: regeneration redux. *Cell* **120**, 295–298 (2005).
95. Laugwitz, K. L. *et al.* Postnatal isl<sup>1</sup> cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433**, 647–653 (2005).
96. Messina, E. *et al.* Isolation and expansion of adult cardiac stem cells from human and murine heart. *Circ. Res.* **95**, 911–921 (2004).
97. Smith, R. R. *et al.* Regenerative potential of cardiosphere-derived cells expanded from percutaneous endomyocardial biopsy specimens. *Circulation* **115**, 896–908 (2007).
98. Bearzi, C. *et al.* Human cardiac stem cells. *Proc. Natl Acad. Sci. USA* **104**, 14068–14073 (2007).
99. Quaini, F. *et al.* Chimerism of the transplanted heart. *N. Engl. J. Med.* **346**, 5–15 (2002).
100. Urbanek, K. *et al.* Intense myocyte formation from cardiac stem cells in human cardiac hypertrophy. *Proc. Natl Acad. Sci. USA* **100**, 10440–10445 (2003).

**Acknowledgements** R.P. and L.W.v.L. were supported by the European Community's Sixth Framework Programme (Heart Repair). ES Cell International provided the human ES3-GFP.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to C.L.M. (c.l.mummery@lumc.nl).



# Tolerance strategies for stem-cell-based therapies

Ann P. Chidgey<sup>1</sup>, Daniel Layton<sup>1</sup>, Alan Trounson<sup>1†</sup> & Richard L. Boyd<sup>1</sup>

**There is much interest in using embryonic stem cells to regenerate tissues and organs. For this approach to succeed, these stem cells or their derivatives must engraft in patients over the long term. Unless a cell transplant is derived from the patient's own cells, however, the cells will be targeted for rejection by the immune system. Although standard methods for suppressing the immune system achieve some success, rejection of the transplant is inevitable. Emerging approaches to address this issue include 're-educating' the immune system to induce tolerance to foreign cells and reducing the immune targeting of the transplant by administering 'self stem cells' instead of foreign cells, but each of these approaches has associated challenges.**

Some of the most debilitating diseases, such as degenerative diseases, cannot be treated by administering single drugs or by technical approaches, such as replacing damaged tissue. This realization has catalysed enormous interest in the potential applications of stem cells. The ability of stem cells to reproduce themselves (self-renew), to form cell lines and, more importantly, to differentiate into many cell types (coupled with the inherent ability of some types of adult stem cell to suppress the immune system) makes them an ideal focus for the treatment of diseases as diverse as inflammatory conditions and cancer, and for regenerative medicine, in which stem cells (or more mature therapeutic derivatives) are used to repair or replace damaged cells or tissues. Embryonic stem (ES) cell lines, which are derived from the inner cell mass of the blastocyst (Fig. 1), have attracted the most attention because of their plasticity, potentially being capable of forming any cell or tissue in the body. The number of protocols available for the directed differentiation of ES cells into cells that can be transplanted for therapy is rapidly increasing. Cells of various tissues can now be generated, including neurons and glia, cardiac muscle cells, blood progenitor cells, hepatocytes, retinal precursor cells, lung epithelial cells and  $\beta$ -cells of pancreatic islets<sup>1–3</sup>. But there are important ethical and safety issues that need to be addressed before cells derived from ES cells can be used in the clinic. Crucially, unless the transplanted cells are autologous (derived from the patient), they will be confronted by a hostile immune system, as would any transplant from a foreign (that is, allogeneic, when from human) source, and prolonged periods of immunosuppression might be required so that these cells are not rejected by the body.

The most obvious way to ensure that a patient accepts a stem-cell transplant from a donor over the long term would be to use the patient's own immunological mechanisms for inducing tolerance to 'self' antigens (mechanisms that prevent the immune system from attacking the patient's own proteins, cells, tissues and organs). These mechanisms could be used, in a similar way, to induce tolerance to the donor's antigens present at the surface of transplanted cells. As logical as this seems, however, it is highly conspicuous how infrequently this strategy has been applied clinically. The reason is simple: the central organ of the immune system that establishes and maintains tolerance to self antigens — the thymus — undergoes severe atrophy with age. Hence, the success of inducing tolerance to a foreign transplant in adults will rely

mainly on rejuvenating the thymus to allow uptake of donor cells (for example, haematopoietic stem cells (HSCs)) and therefore long-term presentation of the donor's antigens in the thymus, where the patient's T cells that recognize these antigens can be eliminated. As an additional safe-guard, the donor's antigens can also induce the development of regulatory T ( $T_{reg}$ ) cells in the thymus.  $T_{reg}$  cells can also be produced in the periphery, and these cells function together with those from the thymus to prevent the activation of any donor-reactive T cells that escaped deletion in the thymus.

In addressing such issues of immunological tolerance, clinical attention has recently focused on adult stem cells. If therapies were based on autologous adult stem cells, then they would not result in rejection or be subject to the same ethical considerations as the use of ES cells. A key practical problem with adult stem cells, however, is that they lack multipotency (the ability to form many cell types) and therefore usually need to be sourced from the relevant target tissue. For a particular patient, however, stem cells might not be present in this tissue or might be affected by the disease. Therefore, to reduce the immunological targeting that would occur if foreign cells were transplanted, strategies have been developed to produce 'self' stem cells that are pluripotent: that is, able to form all three germ layers — the endoderm (which becomes the liver and pancreas), the mesoderm (heart muscle) and the ectoderm (neural tissue) — and therefore give rise to all the cell types of the body. These approaches include somatic cell nuclear transfer (in which the nucleus of a cell from the patient is transferred into an enucleated oocyte) and the transduction of adult somatic cells from the patient by using retroviral constructs containing specific genes associated with pluripotentiality, generating cells known as induced pluripotent stem (iPS) cells<sup>4–6</sup> (Fig. 1).

In this review, we discuss the problem of immunological rejection for translating stem-cell-based therapies into the clinic, and new approaches to overcoming this problem.

## Immunoprivileged versus immunogenic stem cells

Some stem cells seem to be 'invisible' to the host's immune system, whereas others are perceived as completely foreign and will be rejected. What are the defining limits of these alternatives, and are they specific to the stem-cell type?

<sup>1</sup>Monash Immunology and Stem Cell Laboratories, Building 75, STRIP, Monash University, Wellington Road, Clayton 3800, Victoria, Australia. <sup>†</sup>Present address: California Institute for Regenerative Medicine, 210 King Street, San Francisco, California 94107, USA.



## ES cells and their derivatives

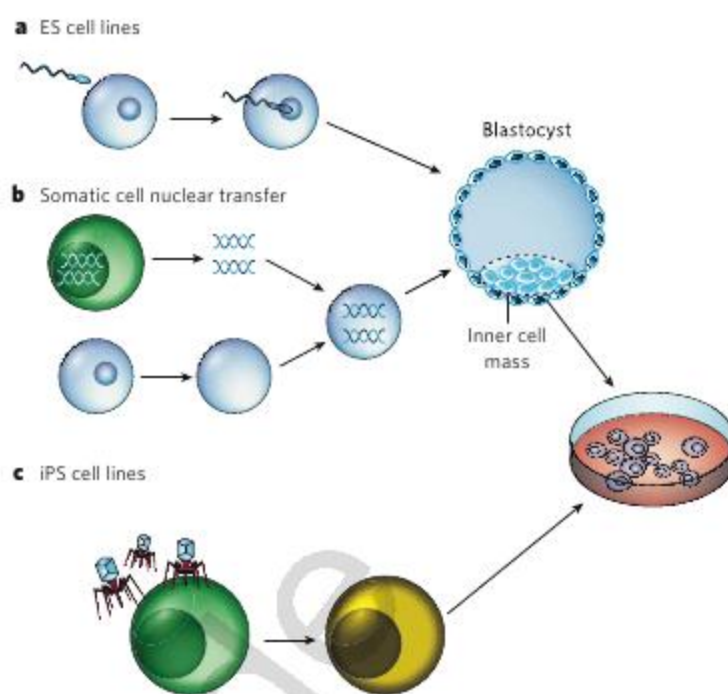
One of the main concerns about transplanting human ES cells is the possibility of immunological rejection (see ref. 7 for a review). Cells targeted by the immune system usually express major histocompatibility complex (MHC) molecules (known as human leukocyte antigen (HLA) molecules in humans and H-2 molecules in mice) or minor histocompatibility (H) antigens. In addition, the ABO blood-group antigens and related blood-group antigens are highly immunogenic (that is, elicit strong immune responses), but they are less likely to be of concern during ES-cell transplantation, because ES cell lines can be selected for the universal donor phenotype (blood group O). ES cells might also express embryo-specific antigens, which would be recognized as foreign by the adult immune system.

HLA class I molecules are present at the surface of almost all nucleated cells, whereas HLA class II molecules are typically restricted to antigen-presenting cells (APCs, the most important of which are dendritic cells), as well as other immune cells such as B cells and a subset of T cells. Both classes of protein can initiate an immune response through the presentation of antigen to T cells. H antigens, although less immunogenic than HLA or ABO antigens, can also induce rejection<sup>8</sup>. Human ES cells produce relatively small amounts of HLA class I molecules and no HLA class II molecules, but some produce H antigens<sup>9</sup>. However, the expression of HLA class I molecules would be expected to increase if ES-cell differentiation follows normal developmental pathways, rendering them susceptible to immunological rejection. Indeed, HLA class I expression is upregulated in embryoid bodies<sup>9</sup>, which are three-dimensional aggregates of ES cells that have begun to differentiate. It is further increased in teratomas<sup>9</sup>, which are tumours that contain heterogeneous tissues and are formed (in this case) by implantation of undifferentiated ES cells. Exposure to soluble interferon- $\gamma$ , which is often present during infections and inflammatory conditions, also increases HLA class I expression, not only by ES cells but also by their derivatives<sup>9</sup>. By contrast, HLA class II expression cannot be induced in this way<sup>9</sup>, and co-stimulatory molecules that are required for the initiation of an immune response (through HLA molecules) are also not expressed, consistent with the inability of ES cells and their derivatives to present antigen to T cells unless the ES cells differentiate into cells of a haematopoietic lineage<sup>10,11</sup>. It has been suggested that a human ES cell bank containing about 150 human ES cell lines that express various combinations of HLA molecules would encompass enough variation to match the HLA haplotypes of most of the population; the appropriate cells could then be administered together with a mild immunosuppressive regimen<sup>12</sup>. However, even a single difference in H antigen between the donor and the recipient is enough to facilitate rejection of a 'graft'<sup>13</sup>.

There is also the potential for the patient's immune system to recognize and respond to fetal antigens or other antigens expressed by ES cells, because these antigens are not present in the thymus and therefore are not accepted as self antigens. Fetal antigens could be present at the surface of ES cells, leading to direct immune attack, or they could be shed by ES cells and captured by the patient's APCs, which would present them to the immune system in a manner similar to other foreign antigens. In either case, the ES cells would be rejected (Fig. 2). Therefore, any endogenous immunosuppressive properties of ES cells, or therapeutic cells derived from them, cannot be relied on to enable these cells to be accepted by the patient, so specific strategies to induce tolerance to an ES-cell-based transplant are required.

## Adult stem cells

In contrast to ES cells, autologous adult stem cells are ideal for transplantation because they are not rejected. Placenta-derived stem cells such as human amnion epithelial cells have been shown to have ES-cell-like properties, including the potential to form all three germ layers *in vitro*<sup>14</sup>. Importantly, unlike ES cells, these cells do not form teratomas when transplanted<sup>15</sup>. They are therefore a useful source of stem cells for cell transplantation and for regenerative medicine approaches, free of the controversial ethical issues and the possible safety concerns associated with ES cells. Furthermore, they could be used in an autologous



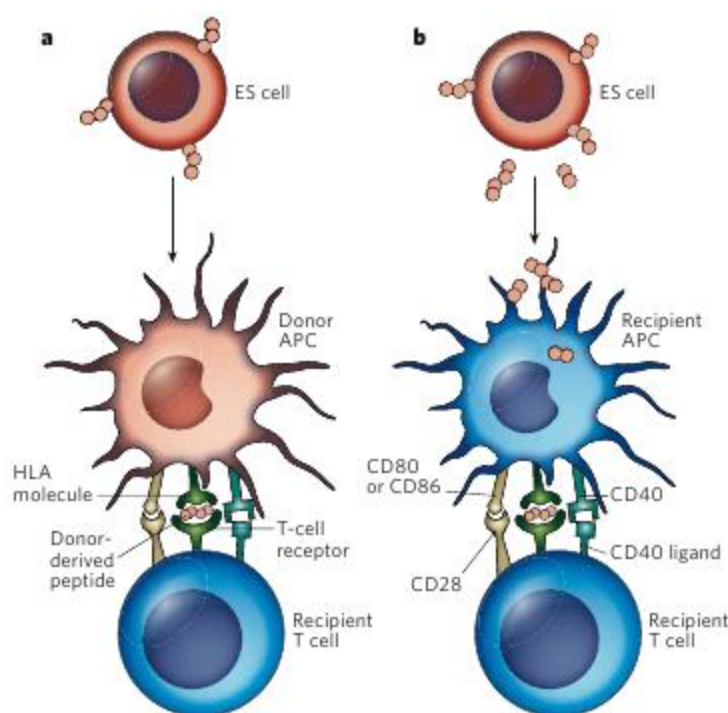
**Figure 1 | Sources of stem cells.** Stem cells can be produced and/or isolated from a variety of sources. **a**, Embryonic stem (ES) cell lines are generated from the fusion of a sperm and oocyte, which is then grown to the blastocyst stage of development. The inner cell mass of the blastocyst is extracted and cultured in specific conditions to maintain the cells in an undifferentiated state. **b**, Somatic cell nuclear transfer involves the removal of the genome of a somatic cell (from any site from an individual of any age) and the removal of the nucleus (enucleation) of an oocyte. The genome of the somatic cell is then transferred into the enucleated oocyte, and the intracellular environment of the oocyte then alters the gene-expression pattern of the somatic cell genome so that the cell becomes a more primitive (less committed) cell type. **c**, Induced pluripotent stem (iPS) cell lines are generated by the dedifferentiation of a somatic cell, by introducing retroviral vectors that carry genes encoding a set of defined factors.

transplant setting if they were collected, cryogenically preserved and stored at birth for later application. In an allogeneic transplant setting (similarly to ES-cell-based transplants), strategies to overcome rejection would be required.

The isolation of tissue-specific stem cells for expansion *in vitro* and transplantation back into the patient is the ideal strategy. However, in most cases of degenerative disease, it is unlikely that enough unaffected stem cells will be isolated, necessitating the use of stem cells from an allogeneic source. Furthermore, stem-cell maintenance and function deteriorate with normal ageing: there are alterations in cell-fate specificity, expansion capacity, telomere length and lifespan, all of which affect whether autologous stem cells can be used (see ref. 16 for a review). Age can also affect the ability of the recipient's stem-cell niches to harbour and support the self-renewal of tissue-specific stem cells, both endogenous stem cells and transplanted ones (see ref. 17 for a review). Thus, enhancing the regenerative potential of aged stem-cell niches before transplantation would be of great clinical benefit.

In terms of their immunogenicity, adult stem cells are more difficult to classify uniformly than ES cells, in part because of the diversity of their sources and the heterogeneity of cell types in each source. In adults, multipotent stem cells known as tissue-specific stem cells reside in a variety of organs and tissues and mediate homeostatic maintenance, repair and regeneration. Transplantation of one type of adult tissue-specific stem cell — neural stem cells — has been studied widely as a potential regenerative therapy, owing to the poor regenerative capacity of the central nervous system. Initially, it was thought that the central nervous system was an immunoprivileged site (that is, a site where normal immune responses to foreign agents do not occur), but it has since been shown that this is not always the case. Neural grafts can be rejected if there is a local inflammatory condition that facilitates the





**Figure 2 | Direct and indirect rejection of a transplant.** Allogeneic transplants can be rejected directly or indirectly. **a**, Direct rejection involves either APCs that were transplanted from the donor or their precursors that have differentiated into APCs after transplantation. When a donor APC comes into contact with a recipient T cell that recognizes one of the donor APC's HLA molecules and/or donor-derived antigen, this APC initiates activation of the recipient T cell through the ligation of co-stimulatory molecules (such as CD80 or CD86 with CD28, and CD40 with CD40 ligand). **b**, Indirect rejection occurs through the shedding of antigen from transplanted cells or tissues. These antigens are taken up by recipient APCs, which then present the foreign antigen to recipient T cells. This can result in the activation of donor-reactive recipient T cells, which then destroy the graft.

entry of lymphocytes<sup>18,19</sup>. Although neural stem cells express only small amounts of HLA class I molecules, HLA class I expression is upregulated in response to pro-inflammatory signals and when these cells differentiate into neurons<sup>20</sup>. Similarly, muscle progenitor cells, which have shown much promise for the treatment of patients with degenerative muscle diseases such as muscular dystrophy, express HLA class I molecules and can be induced to express HLA class II molecules<sup>21,22</sup>. Accordingly, transplantation of muscle stem cells was of limited therapeutic value in clinical trials<sup>23,24</sup>, a failure that was later attributed, in part, to rapid immunological rejection of the transplant in the absence of immunosuppression<sup>24</sup>. Therefore, in an allogeneic setting, the engraftment of adult tissue-specific stem cells will almost certainly require strategies to induce immunological tolerance to the cells and thus overcome the problem of rejection.

One type of adult stem cell that might be immunoprivileged, as a result of endogenous immunosuppressive properties, is the mesenchymal stem cell (MSC). MSCs were first isolated from the bone marrow as an adherent fibroblast-like population<sup>25</sup> and have now been identified in many tissues, including umbilical-cord blood<sup>26</sup>, the placenta<sup>27</sup>, fat<sup>28</sup> and amniotic fluid<sup>29</sup>. They are defined by the presence or absence of specific cell-surface molecules, their self-renewal capacity and their ability to form bone, fat and cartilage (see ref. 30 for a review). Using autologous MSCs to restructure damaged bone has had some clinical success<sup>31</sup>, but it has not yet been clearly shown whether MSCs durably engraft in an allogeneic host. There is, however, a rapidly growing body of literature on using MSCs to treat individuals with diseases for which anti-inflammatory and/or regenerative outcomes are beneficial<sup>32</sup>. Transplantation of MSCs clearly benefits some patients, but the effects are equivocal in many others. A more exhaustive clinical evaluation is required to define accurately the optimal cell type, cell dose, conditioning of the patient before the transplant, subsequent treatment regimen and safety requirements for a successful MSC-based

transplant; many phase I, II and III trials are in progress to address these issues. To optimize the clinical application of MSCs, it will also be crucial to determine the mechanisms underlying their engraftment.

### Avoiding rejection through 'natural' immunosuppression

MSCs have also attracted interest because of the possibility of harnessing their immunosuppressive properties. *In vitro*, MSCs inhibit the activation and proliferation of T cells, which are mediated, at least in part, by various soluble factors, including transforming growth factor- $\beta$ , prostaglandin E<sub>2</sub> and indoleamine 2,3-dioxygenase<sup>33</sup>. Even though MSCs express HLA class I molecules<sup>34</sup>, and even though HLA class II expression can be induced with interferon- $\gamma$ , MSCs do not initiate an allogeneic response *in vitro*<sup>34,35</sup>, and they are not rapidly rejected *in vivo*. However, the clinical potential of these immunosuppressive properties is a matter of speculation, because most data have been obtained from *in vitro* studies. If MSCs are shown to be immunosuppressive in controlled *in vivo* settings, then transplantation of MSCs together with an allogeneic graft (allograft) might facilitate the acceptance of stem-cell-based therapies, including ES-cell derivatives.

### Enhancing self-tolerance for acceptance of stem-cell therapies

Immunological self-tolerance has two facets: central (or thymic) tolerance (Box 1) and peripheral (or non-thymic) tolerance. Central tolerance involves the negative selection of T cells (that is, the intrathymic deletion of T cells that display an antigen receptor with a high affinity for self antigens) and the generation of T<sub>reg</sub> cells. Peripheral tolerance is a consequence of inappropriate co-stimulation after T cells have exited the thymus, which results in T-cell anergy (that is, non-responsiveness) or T<sub>reg</sub>-cell induction, instead of T-cell activation<sup>36</sup>.

Both central tolerance and peripheral tolerance are logical targets for inducing the acceptance of foreign grafts, including stem-cell-based therapies. Manipulating central tolerance would be an ideal approach: if a 'donor tolerant' immune system could be generated by physically deleting the relevant donor-reactive T cells intrathymically (before the T cells exit to the blood), then this would avoid the need for the long-term, non-specific immunosuppressive regimen that is currently used to suppress peripheral T cells. However, central tolerance depends on two fundamental conditions: that antigen (in this case, donor-derived antigen) can enter the thymus for presentation to, and deletion of, developing T cells that recognize the antigen; and that the thymus is functioning sufficiently to generate a new repertoire of T cells. The issue of antigen access to the thymus can be overcome to some extent through transplanting donor HSCs or ES-cell-derived HSCs<sup>37</sup> (discussed below), but addressing thymic function is problematic. It is well recognized, yet paradoxical (given the continued need for our immune system throughout adulthood), that the thymus becomes severely atrophied early in life, a process that is clearly evident by puberty (concomitant with the increase in production of sex steroids) (see ref. 38 for a review). Given that most degenerative conditions occur later in life, a major consideration when attempting to induce tolerance to transplants is the need to reactivate the thymus in adult patients to facilitate its functional involvement. As an adjunct to improving central tolerance, strategies for manipulating peripheral tolerance to ensure complete prevention of allograft rejection should also be considered. Both approaches are discussed in this section.

### Introducing donor-derived antigen into the thymus

HSCs or their immediate progenitors have preferential access to the thymus, so the induction of tolerance through HSC transplantation (or bone-marrow transplantation) is the subject of intense investigation. Early studies by Suzanne Ildstad and colleagues<sup>39,40</sup> showed that transplantation of HSCs into young animals, in conjunction with depletion or ablation of the recipient's immune system, led to engraftment of the HSCs and ultimately to the coexistence of donor and recipient haematopoietic cells (known as mixed chimaerism). After HSC engraftment, APCs of donor origin migrate to, or (more probably) develop within, the thymus and contribute to the negative selection of donor-reactive T cells, resulting in tolerance to that donor. In addition, T<sub>reg</sub> cells of



donor origin can be generated in the thymus, and they then populate the periphery, contributing to peripheral-tolerance mechanisms. Furthermore, in rats, ES-cell-like cells have been shown to induce tolerance after transplantation, but only in animals with a competent thymus<sup>41</sup>, providing important proof of principle that ES-cell-based therapies are possible if there is sufficient thymic function. In these animals, tolerance is probably induced by the intrathymic derivation of dendritic cells or their progenitors from the ES cells.

Another strategy would be to co-transplant ES-cell-derived HSCs or APCs<sup>42</sup> with therapeutic cellular products derived from the same ES cells (for example, neural cells) (Fig. 3), to establish bone-marrow and thymic chimaeras. In theory, these ES-cell-derived APCs (such as HSC-derived dendritic cells, either derived *in vivo* or developed *in vitro*) would migrate to the thymus and present donor-derived antigens to developing T cells, rendering the recipient tolerant to the ES-cell derivatives. Richard Burt *et al.*<sup>43</sup> have shown that known differentiation factors can promote ES cell lines to differentiate into HSCs. The engraftment of these ES-cell-derived HSCs in allogeneic recipients was not associated with signs of graft-versus-host disease, a state in which donor T cells can attack host tissue. Furthermore, using *in vitro* assays, bidirectional T-cell tolerance was demonstrated. Interestingly, transplantation of ES-cell-derived HSCs could also ameliorate autoimmune disease in an immunodeficient mouse model of diabetes<sup>44</sup>. It is not known whether secondary cellular or tissue allografts (that is, ES-cell-derived cells or tissues, such as neural cells or pancreatic islet cells) would be accepted, but *in vitro* data suggest that they would. Strategies for allogeneic adult stem-cell engraftment could follow the same principle as this ES-cell-based approach: transplanting donor HSCs to induce central tolerance to tissue-specific stem cells from the same donor.

Recently, clinical proof of principle was obtained for this strategy to induce central tolerance. A girl of 9 years of age (and therefore with robust thymic function) was given a liver allograft and acquired complete haematopoietic chimaerism and tolerance<sup>45</sup>. Presumably, enough HSCs were fortuitously present in the liver to enable their migration to, and engraftment in, the bone marrow and full chimaerism between donor and recipient after thymic uptake of HSC derivatives. The subsequent production of donor dendritic cells that present donor-derived antigen to developing T cells in a competent pre-pubertal thymus presumably led to complete tolerance to the donor tissue. The importance of a functional bone marrow and thymus is crucial for the success of this strategy, as they are needed to provide an environment conducive to the engraftment and development of HSCs.

The stem-cell niche in the bone marrow is a complex and specialized microenvironment that provides the signals and growth factors that are required for HSC maintenance, controlled division, lineage commitment, haematopoietic and lymphoid lineage differentiation, and migration. Any compromise to this microenvironment as a result of ageing or cytoablative preconditioning regimens directly affects the trans-marrow migration of transplanted donor HSCs, their attachment to the stem-cell niche and, subsequently, their development. Therefore, the ideal strategy would be to optimize the conditions for successful engraftment, particularly for patients of more than 50 years of age, before transplantation. For example, one of the crucial components of the stem-cell niche in the bone marrow, the osteoblast population (which lines the endosteum), has been a target for therapeutic manipulation. In animal models, pre-treatment with human parathyroid hormone led to an enlargement of the HSC pool, as well as protection against damage induced by cytoablative preconditioning by chemotherapy and an improvement in engraftment of the bone-marrow transplant<sup>46</sup>, offering the possibility of improving chimaerism with donor cells in ageing patients.

### Restoring the ageing thymus for tolerance induction

The human thymus remains functional, albeit highly compromised, throughout adulthood, but it slowly involutes from as early as 2 years of age, with accelerated atrophy associated with puberty. By 40–50 years of age, it generally has less than 10% of its maximal cellular volume<sup>47</sup>. This correlates with a reduction in the export of newly derived naive

### Box 1 | The thymus and its role in transplant immunology

One of the crucial requirements for the induction of tolerance to self antigens is having a fully functional thymus that receives haematopoietic stem cells (HSCs) and progenitor cells committed to the lymphoid lineage from the bone marrow.

The thymus comprises two functionally distinct areas — the cortex and the medulla — that each contains distinct stromal cells known as thymic epithelial cells (TECs). Progenitor cells from the bone marrow enter the thymus at the cortico-medullary junction and undergo a series of differentiation and proliferation events as they interact with, and migrate in a directed manner through, the intrathymic niches in the cortex. In this location, after the cells have differentiated into T cells, they undergo lineage restriction (that is, commitment to the CD4<sup>+</sup> or CD8<sup>+</sup> T-cell lineage) and positive selection, a process in which the only cells selected to survive are those that produce a functional T-cell receptor that interacts with low avidity to self peptide presented by cortical TECs (cTECs).

The selected CD4<sup>+</sup> and CD8<sup>+</sup> T cells continue to migrate through to the medulla, where they undergo negative selection and further maturation events before exiting into the periphery as mature, self-tolerant T cells. Negative selection is mainly mediated by HSC-derived dendritic cells, which present self peptides in the context of MHC molecules to developing T cells as they migrate into the medulla. This process results in the deletion of T cells that bind with high avidity (see ref. 81 for a review), thus eliminating any potentially autoreactive T cells. However, not all self antigens are presented by these dendritic cells. Thus, in addition, under the direction of the transcription factor AIRE, specialized mature (CD80<sup>+</sup>) medullary TECs (mTECs) further purge the emerging naive T-cell repertoire of autoreactive cells. They do so by expressing and then presenting numerous self antigens that are found in the periphery but not in the thymus, such as tissue-specific and sequestered self antigens (including myelin components, pro-insulin and tissue-specific collagens)<sup>82,83</sup>. These AIRE-producing cells are only a minor subset of the mTEC population; they produce large amounts of MHC class II molecules (denoted mTEC<sup>hi</sup>)<sup>84</sup> and are mature cells derived from mTEC<sup>low</sup>CD80<sup>+</sup> cells<sup>85</sup>. More recently, thymic CD4<sup>+</sup>CD3<sup>+</sup> cells (known as lymphoid-tissue inducer cells) have been found to control the development of AIRE-producing mTECs specifically, through signals from receptor activator of nuclear factor- $\kappa$ B (RANK) ligand<sup>86</sup>. The thymus also produces regulatory T cells, which can suppress self-reactive T cells that escape deletion in the thymus<sup>86,87</sup>.

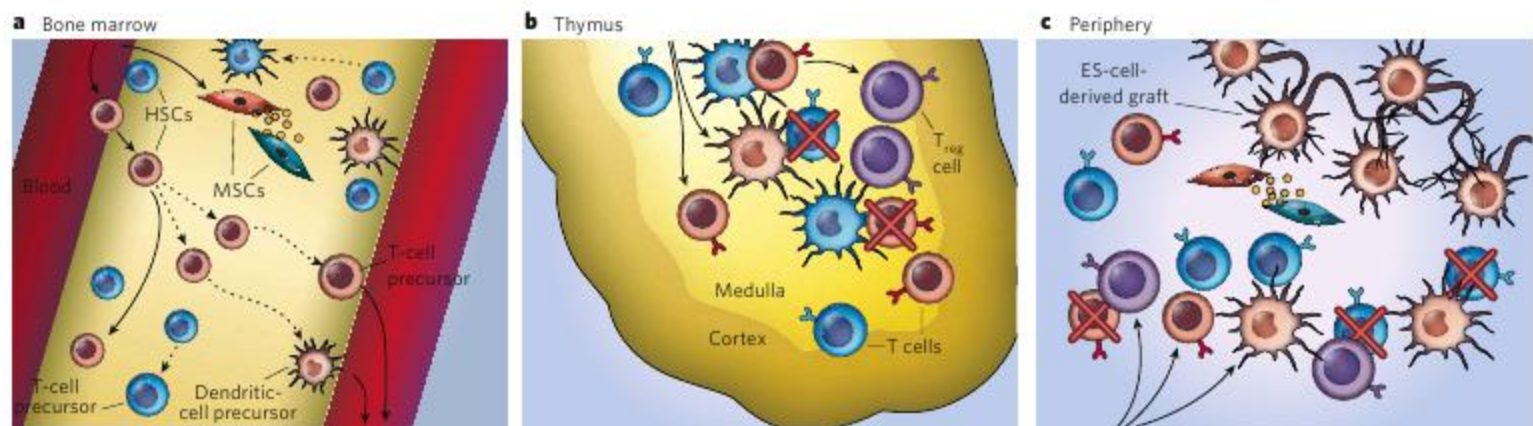
The outcome of these complementary tolerogenic mechanisms is the maintenance of self-tolerance and the prevention of autoimmunity. The optimal way to induce tolerance to donor cells would be to establish long-term chimaerism of leukocytes (derived from both donor HSCs and host HSCs) (Fig. 3). However, the thymus atrophies with age, and it is damaged by cytoablative conditioning regimens (which are commonly administered before a transplant, to create 'space' for the engraftment of donor HSCs). These factors might therefore preclude efficient functioning of the thymus, thereby preventing the development of adequate donor-host chimaerism and subsequent long-term tolerance to the donor in adults. These patients would therefore need to receive immunosuppressive agents indefinitely.

T cells and therefore imposes considerable limitations on any tolerance-induction strategy for use in adult patients.

Restoration of thymic function is therefore a major clinical challenge. At present, several approaches are in preclinical studies or clinical trials. The first approach is to target the endocrine-immune axis by using analogues of luteinizing-hormone-releasing hormone (LHRH) to inhibit pituitary gonadotropin and sex-steroid production in adults. The second approach is to administer thymic growth factors that are known to decline in production with age. These factors include growth hormone, interleukin-7 (IL-7) and factors that induce cell-type-specific proliferation, such as keratinocyte growth factor (KGF, also known as FGF7, which stimulates thymic epithelial cells (TECs)) (Fig. 4).

In terms of the first approach — manipulating the endocrine-immune axis — we and others have shown that removal of sex steroids





**Figure 3 | Generating mixed haematopoietic chimaerism to induce tolerance.** Immunological tolerance can be induced to a regenerative therapy, such as a transplant, by generating mixed haematopoietic chimaerism. Donor cells are shown in red, and recipient cells are shown in blue. Solid arrows indicate movement; dashed arrows indicate that multiple differentiation events are involved. **a**, HSCs from an allogeneic donor will engraft in the bone marrow if suitable preconditioning has been carried out. The co-transplantation of MSCs can aid in the development of a graft-permissive environment, through cell–cell contact and/or the secretion of immunosuppressive molecules. After the HSCs have engrafted, they can produce T-cell precursors and dendritic-cell precursors, which then migrate to the thymus. **b**, The T-cell precursors enter the thymus through blood vessels at the cortico–medullary junction and migrate through the cortex, where they undergo positive selection (not shown). They then

migrate to the medulla, where donor and recipient dendritic cells can delete any donor-reactive or recipient-reactive T cells. This process, known as negative selection, ensures that the T cells that exit the thymus should not recognize and respond to either donor or recipient cells. T<sub>reg</sub> cells are also produced in the medulla. The tolerized T cells, together with donor-specific and recipient-specific T<sub>reg</sub> cells, subsequently migrate to the periphery. **c**, In the periphery, the engrafted HSC-derived donor dendritic cells (if they are tolerogenic) can initiate the deletion of donor-reactive T cells that have escaped thymic selection, and the T<sub>reg</sub> cells can induce anergy in donor-reactive T cells. Both of these processes promote tolerance to donor cells and acceptance of the graft. MSCs can also aid in the prevention of graft rejection by dampening the inflammatory response elicited by any T cells that become activated at the graft site, through cell–cell contact and/or the secretion of immunosuppressive molecules.

markedly rejuvenates the thymus and increases bone-marrow cellularity and function within 2–4 weeks in aged mice<sup>48–50</sup> and after chemotherapy-induced damage<sup>51</sup>. These improvements were associated with an increase in the production of IL-7-responsive progenitor cells in the bone marrow and, subsequently, an increase in the export of B cells, as well as in an increase in the number of thymus-derived naive T cells in the blood, overall contributing to an increase in immune competence. In addition, regeneration and normalization of thymic stromal cell ratios (TEC number to fibroblast number) and thymic architecture (both of which are disrupted by ageing), as well as better engraftment in either an autologous or allogeneic bone-marrow transplant setting<sup>52,53</sup>, reflected an improvement in the quality of both the thymic microenvironment and the bone-marrow microenvironment.

Reversible inhibition of sex-steroid production can be achieved in humans by using an LHRH agonist (LHRH-A); this agent has been used without complications for more than 25 years, for treating prostate and breast cancers, precocious puberty and endometriosis. Closely in line with our mouse studies<sup>50</sup>, we found that patients with prostate cancer who were treated with LHRH-A have increased numbers of naive T cells (both CD4<sup>+</sup> and CD8<sup>+</sup>) that contain T-cell-receptor excision circles (a marker of recent thymic emigrants), reflecting improved thymic function<sup>50</sup>. Furthermore, in a pilot phase II clinical trial, treatment with LHRH-A similarly increased the generation of naive T cells after either autologous or allogeneic bone-marrow transplantation, and improved HSC engraftment<sup>54</sup>.

In terms of the second strategy — restoring the thymus by administering thymic growth factors — preclinical studies have shown that KGF, growth hormone, IL-7 and FLT3 ligand can augment T-cell (and in some cases B-cell) survival and production. KGF, which is produced by mesenchymal cells, binds specifically to the IIIB isoform of fibroblast growth factor receptor 2, which is present at the surface of TECs. In preclinical studies, KGF was found to restore thymic function in aged rodents<sup>55</sup> and to protect TECs against damage by cytotoxic therapy and graft-versus-host disease<sup>56,57</sup>. In addition, when numerous doses were administered to rhesus macaques in a model of HSC transplantation, KGF increased the reconstitution of the thymus-dependent T-cell population after transplantation of autologous CD34<sup>+</sup> peripheral-blood progenitor cells<sup>58</sup>. Although KGF has now been approved for the treatment

of patients with severe mucositis caused by high-dose chemotherapy, its clinical significance in the context of T-cell reconstitution has yet to be fully demonstrated, but promising findings are emerging.

Growth hormone has been used with limited success to treat patients infected with HIV, resulting in an increase in the mass of thymic tissue and in the number of circulating naive CD4<sup>+</sup> T cells. However, to achieve this, daily injections were required for 6–12 months<sup>59</sup>, and these were associated with severe side effects, particularly the risk of developing diabetes. It might therefore be practical to limit treatment with growth hormone to use in combination with other therapies for the regeneration of immunological niches and the production of lymphoid lineage cells.

Administration of either IL-7 (ref. 60) or FLT3 ligand<sup>61</sup> has been shown to increase the production of lymphoid progenitor cells in the short term after transplantation of bone marrow into immunodeficient mice. In addition, IL-7 resulted in an increase in the number of naive T cells generated<sup>62</sup>. If these results are reproduced in the clinic, then IL-7 might also prove to be an important adjunct to methods for thymic rejuvenation, such as sex-steroid ablation and administration of other growth factors.

### Manipulating peripheral tolerance

Prevention of allograft rejection can be partly achieved in the periphery through blockade of co-stimulatory molecule stimulation and through generation of tolerogenic (that is, tolerance-inducing) dendritic cells, which suppress immune function. In an allogeneic setting, tolerogenic dendritic cells can prolong allograft survival (see ref. 63 for a review). Dendritic cells cultured in the presence of MSCs were unable to upregulate production of the HLA class I molecule HLA-DR or of co-stimulatory molecules, and they showed reduced production of IL-12 and increased production of IL-10 (refs 64, 65). Hence, MSCs seem capable of inducing dendritic cells to enter an immature state or a tolerogenic state, which might be an indirect mechanism of inducing T-cell hyporesponsiveness.

T<sub>reg</sub> cells are also a crucial component of peripheral tolerance, because they can effectively suppress T-cell activation. A lack of T<sub>reg</sub> cells or the presence of defective T<sub>reg</sub> cells is strongly associated with the development of autoimmunity<sup>66</sup>. There are several subsets of T<sub>reg</sub> cells, the two most studied being CD4<sup>+</sup>CD25<sup>+</sup>FOXP3<sup>+</sup> cells, which are naturally



occurring, and  $CD4^+CD25^-FOXP3^+$  cells, which are induced. Induced  $T_{reg}$  cells<sup>67</sup> exert their suppressive effects mainly through the production of the immunosuppressive molecules IL-10 and transforming growth factor- $\beta$ .  $T_{reg}$  cells have been linked to the induction of transplantation tolerance in preclinical studies<sup>68</sup> and to stable renal-allograft function in a separate study<sup>37</sup>. In that study, four patients received, from a relative, a combined kidney and bone-marrow transplant with a single mismatch in the HLA haplotype<sup>37</sup>. In both studies, transient chimaerism was observed in the bone marrow, and this might have led to the elimination of donor-reactive T cells in the thymus through the development of donor dendritic cells from migrating precursors, thereby accounting for the observed loss of *in vitro* activity of the recipient T cells towards the donor cells. Although the exact mechanism is unclear, a subsequent switch from deletion of T cells in the thymus to peripheral tolerance seems to be important, with large numbers of  $FOXP3^+T_{reg}$  cells being found in the renal allografts.

Dendritic cells can also induce the generation of  $T_{reg}$  cells (see ref. 69 for a review): allogeneic mouse dendritic cells, in combination with IL-2, induced the clonal expansion of  $FOXP3^+T_{reg}$  cells<sup>70</sup> specific for that strain of mice. In theory, the production of dendritic cells could therefore be induced from ES cells, and these cells could then be used to generate  $T_{reg}$  cells specific for those ES cells and their derivatives<sup>71</sup>. This process could occur *in vivo* after transplantation of ES cells or by transplantation of  $T_{reg}$  cells after *in vitro* generation<sup>72</sup>.

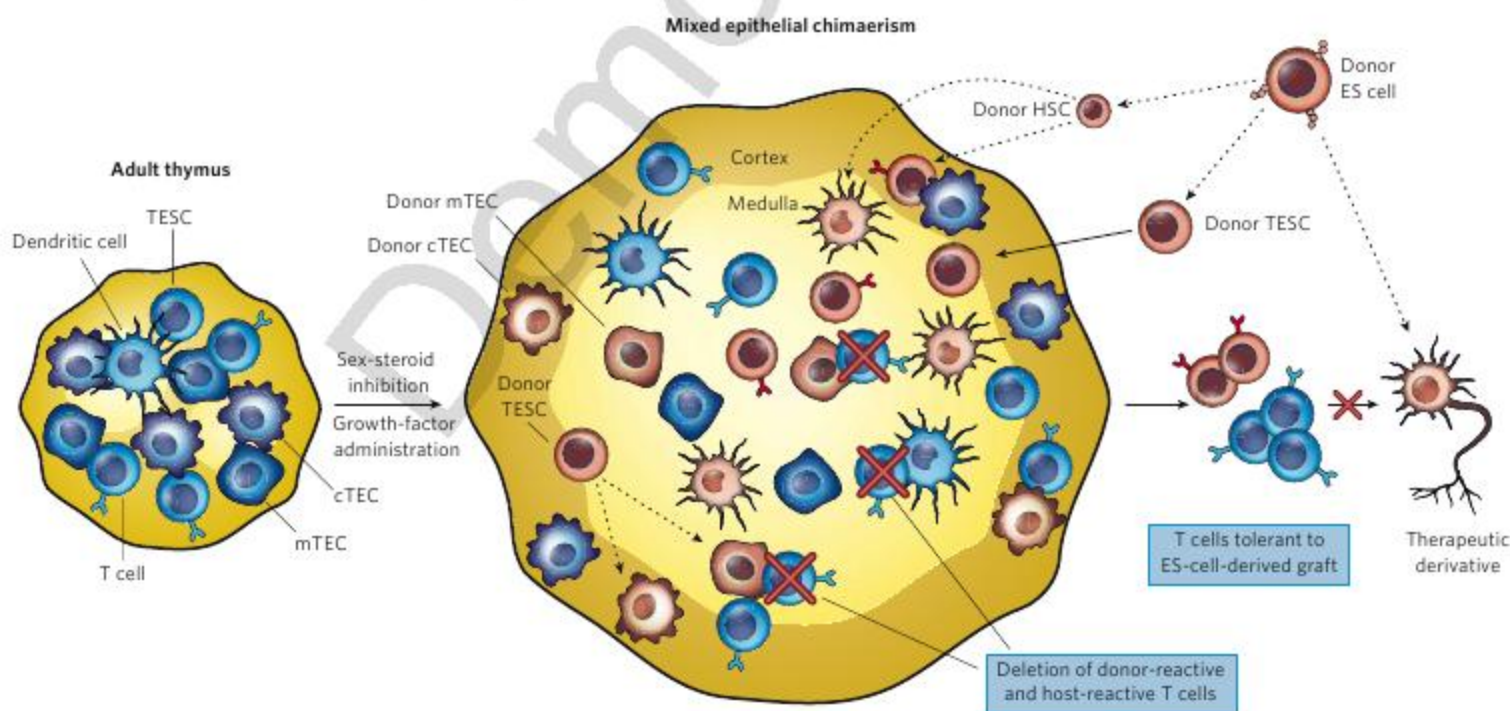
### Using self stem cells to reduce immune targeting of grafts

An alternative to manipulating self-tolerance mechanisms to avoid rejection of allogeneic stem-cell-derived grafts has gained attention recently: the possibility of dedifferentiating mature autologous cells. Somatic cell nuclear transfer (also known as therapeutic cloning) and the generation of iPS cells (Fig. 1) can provide stem cells that, in theory, have an almost identical antigenic complement to the recipient.

Somatic cell nuclear transfer uses the cytoplasmic environment of an enucleated oocyte to reprogramme the donor genome extracted from a somatic cell. As is the case for ES cells, the oocyte that contains the

donor genome is grown to the blastocyst stage, and the inner cell mass is removed and cultured to form ES cell lines. So far, genuine ES cells have been developed using this technique only for mice and rhesus macaques, but blastocyst-stage embryos of several species, including humans, have been produced<sup>73–75</sup>. At least some ES cell lines developed by somatic cell nuclear transfer do not seem to be immunogenic when transplanted into the donor of the nucleus, even though mitochondria are heterogeneous (with their own source of DNA) and therefore could, in principle, confer immunological differences on the ES cells, rendering them subject to immunological rejection (see ref. 75 for a review). The technical difficulties and the associated controversy over the use of oocytes, however, mean that this approach might be many years away from the clinic.

More recently, surprising progress has been made in reprogramming adult cells to show pluripotent properties similar to those of ES cells. These cells, known as iPS cells, hold considerable promise as an individualized form of therapy with 'self' stem cells. Retroviral transduction of the genes encoding several transcription factors — SOX2, OCT4, KLF4 and Myc — induced the transformation of both embryonic mouse fibroblasts and adult mouse fibroblasts into pluripotent cells that resemble ES cells in morphology, growth patterns and gene expression<sup>5</sup>. The resultant cells were able to form embryoid bodies and teratomas and to contribute to the formation of chimaeric mouse embryos. Despite these functional similarities, however, there are differences in the gene-expression patterns of ES cells and iPS cells. After this initial proof of principle, iPS cells were produced on the basis of selecting cells that expressed the genes encoding NANOG and/or OCT4 after retroviral transduction, which is in contrast to the original study, in which cells were selected only on the basis of expression of the gene encoding FBX15 (a target of OCT4)<sup>4,6</sup>. These iPS cells had stable endogenous expression of SOX2 and OCT4, in addition to other pluripotency markers, and were again similar to ES cells in terms of gene expression. This method of iPS-cell production has also been applied to human fetal, neonatal and adult cells<sup>4,76–78</sup>. Whether this process will render cells susceptible to immune attack, as a result of their altered gene expression, has yet to be determined.



**Figure 4 | Generating mixed epithelial chimaerism to induce tolerance.** After puberty, the thymus becomes markedly involuted, resulting in indistinct cortical and medullary areas and a corresponding reduction in thymic function. Reversal of this ageing-induced atrophy — through factors such as LHRH-A, KGF and growth hormone — can contribute to the regeneration of the thymus to a state resembling its prepubescent size and function. In combination with these therapies, thymic epithelial stem cells (TESCs) derived from ES cells, together with HSCs derived from ES cells, could be administered by intrathymic injection and intravenous injection,

respectively. In the rejuvenating thymic environment, the TSCs can differentiate into both cTECs and mTECs, and the HSCs can differentiate into dendritic cells and T-cell precursors. The cTECs can positively select T cells with a functional T-cell receptor, and the mTECs, together with the dendritic cells, can negatively select T cells that react to recipient and donor cells, resulting in the emigration of tolerant T cells that are not reactive to self antigens or to a transplant from the donor. Donor cells are shown in red, and recipient cells are shown in blue. Solid arrows indicate movement; dashed arrows indicate that multiple differentiation events are involved.



One of the problems associated with iPS cells is the use of retroviral transduction and the high frequency of tumour formation associated with the introduction of *Myc*<sup>4</sup>. These issues have been partly resolved by removing *Myc* from the production of mouse and human iPS cells and by increasing the selection times<sup>79</sup>. The iPS cells produced in this way were able to produce chimaeras with reduced tumorigenicity in mice; however, omission of *Myc* from the protocol considerably reduced the capacity of human iPS cells to form fibroblasts. Alternative genes — *NANOG* and *LIN28*, instead of *KLF4* and *Myc* — were subsequently introduced to produce human iPS cells; however, it remains to be seen how effective, practical and safe this technique will be<sup>80</sup>.

## Conclusions

Most degenerative or autoimmune diseases are caused by several factors, so the optimal therapy is likely to consist of a combination of treatments, including modulation of the immune system to facilitate allogeneic cell-based therapies (which are exemplified by those based on stem cells). This is certainly the case for many degenerative diseases. Treating these diseases will require transplantation of tissue-specific stem cells or their immediate derivatives, together with initial retardation of the immune response to prevent rejection and rebuilding of a donor-tolerant immune system.

ES cells offer enormous potential for regenerative medicine approaches, but the ethical controversies surrounding the use of these cells are driving the search for alternative sources of stem cells, such as adult tissue-specific stem cells, placenta-derived stem cells and, more recently, fully differentiated somatic cells that are reprogrammed. Many approaches are also being developed to promote the maintenance or the differentiation of tissue-specific stem cells. These include bioengineering of tissues (such as the development of 'organoids', by using biomaterials as scaffolds on which cells can develop in a three-dimensional manner) and inducing stem cells to differentiate by using microRNAs (that is, targeted non-coding RNAs that post-transcriptionally regulate gene expression).

One way to manipulate tolerance mechanisms, in addition to the approaches discussed here, might be to regenerate the atrophied thymus of the recipient and then to inject intrathymically both thymic epithelial stem cells (TESCs) and HSCs derived from the same ES cells (Fig. 4). Alternatively, the ES-cell-derived HSCs could be injected intravenously to facilitate engraftment in the bone marrow and subsequently continual production of T-cell precursors, aiding long-term thymic reconstitution. Another possibility is to create an artificial thymus *ex vivo* using donor TESCs and HSCs chimaeric with host HSCs; the cells would be encompassed in clearly defined biomaterials that contain specific mesenchymal growth factors and lymphoid growth factors to promote TEC production and T-cell production, respectively.

Although iPS cells have the potential to produce an autologous source of pluripotent stem cells for regenerative therapies, many hurdles need to be overcome before the appropriate regulatory approval will be granted for their use in the clinic. Alternative non-viral delivery systems need to be developed to replace the use of an integrating vector such as a retrovirus, which can cause insertional mutagenesis. Some of the issues that will need to be investigated further are the effect of the differentiation state of the target cell on the efficiency of reprogramming, and whether the transcriptional and epigenetic programmes of the iPS cells have been incorrectly altered to predispose them to malignant transformation. However, if such technical problems and safety issues can be resolved, the potential for iPS cells to produce patient-specific stem cells, even if they are initially used just as diagnostic targets *in vitro*, will be a great complement to other regenerative medicine approaches. The stage is thus set for the emergence of new therapeutic regimens that can overcome transplant rejection, thereby allowing the translation of stem-cell-based therapies from the lab to the clinic. ■

- Goulburn, A. & Trounson, A. In *Cell Therapy* (eds Garcia-Olmo, D., Garcia-Verdugo, J.M., Alemany, J. & Gutiérrez-Fuentes, J.A.) 169–185 (McGraw Hill, Madrid, 2008).
- Hoffman, L. M. & Carpenter, M. K. Characterization and culture of human embryonic stem cells. *Nature Biotechnol.* **23**, 699–708 (2005).

- Trounson, A. The production and directed differentiation of human embryonic stem cells. *Endocr. Rev.* **27**, 208–219 (2006).
- Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007). This study refined previous iPS cell studies and showed that live chimaeras can be generated after transplantation of iPS cells that had been selected on the basis of *Nanog* expression, and it highlighted the importance of avoiding the retroviral introduction of *Myc* in patients.
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006). This study resulted in a major shift in stem-cell research by presenting an alternative to ES cells: 'self stem cells' were induced by transfecting normal adult somatic cells with four ES-cell-related genes, generating cells that are now known as iPS cells.
- Wernig, M. et al. *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
- Drukker, M. & Benvenisty, N. The immunogenicity of human embryonic stem-derived cells. *Trends Biotechnol.* **22**, 136–141 (2004).
- Simpson, E. et al. Minor H antigens: genes and peptides. *Eur. J. Immunogenet.* **28**, 505–513 (2001).
- Drukker, M. et al. Characterization of the expression of MHC proteins in human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **99**, 9864–9869 (2002).
- Grinnemo, K. H. et al. Human embryonic stem cells are immunogenic in allogeneic and xenogeneic settings. *Reprod. Biomed. Online* **13**, 712–724 (2006).
- Li, L. et al. Human embryonic stem cells possess immune-privileged properties. *Stem Cells* **22**, 448–456 (2004).
- Taylor, C. J. et al. Banking on human embryonic stem cells: estimating the number of donor cell lines needed for HLA matching. *Lancet* **366**, 2019–2025 (2005).
- Robertson, N. J. et al. Embryonic stem cell-derived tissues are immunogenic but their inherent immune privilege promotes the induction of tolerance. *Proc. Natl Acad. Sci. USA* **104**, 20920–20925 (2007).
- Miki, T., Lehmann, T., Cai, H., Stolz, D. B. & Strom, S. C. Stem cell characteristics of amniotic epithelial cells. *Stem Cells* **23**, 1549–1559 (2005).
- Ilancheran, S. et al. Stem cells derived from human fetal membranes display multilineage differentiation potential. *Biol. Reprod.* **77**, 577–588 (2007).
- Rando, T. A. Stem cells, ageing and the quest for immortality. *Nature* **441**, 1080–1086 (2006).
- Jones, D. L. & Wagers, A. J. No place like home: anatomy and function of the stem cell niche. *Nature Rev. Mol. Cell Biol.* **9**, 11–21 (2008).
- Mason, D. W. et al. The fate of allogeneic and xenogeneic neuronal tissue transplanted into the third ventricle of rodents. *Neuroscience* **19**, 685–694 (1986).
- Wong, G. H., Bartlett, P. F., Clark-Lewis, I., Baty, F. & Schrader, J. W. Inducible expression of H-2 and Ia antigens on brain cells. *Nature* **310**, 688–691 (1984).
- McLaren, F. H., Svendsen, C. N., Van der Meide, P. & Joly, E. Analysis of neural stem cells by flow cytometry: cellular differentiation modifies patterns of MHC expression. *J. Neuroimmunol.* **112**, 35–46 (2001).
- Bao, S. S., King, N. J. & dos Remedios, C. G. Elevated MHC class I and II antigens in cultured human embryonic myoblasts following stimulation with  $\gamma$ -interferon. *Immunol. Cell Biol.* **68**, 235–241 (1990).
- Mantegazza, R. et al. Modulation of MHC class II antigen expression in human myoblasts after treatment with IFN- $\gamma$ . *Neurology* **41**, 1128–1132 (1991).
- Tremblay, J. P. et al. Results of a triple blind clinical study of myoblast transplantations without immunosuppressive treatment in young boys with Duchenne muscular dystrophy. *Cell Transplant.* **2**, 99–112 (1993).
- Guerette, B., Asselin, L., Vilquin, J. T., Roy, R. & Tremblay, J. P. Lymphocyte infiltration following allo- and xenomyoblast transplantation in mice. *Transplant. Proc.* **26**, 3461–3462 (1994).
- Friedenstein, A. J., Gorskaja, J. F. & Kulagina, N. N. Fibroblast precursors in normal and irradiated mouse hematopoietic organs. *Exp. Hematol.* **4**, 267–274 (1976).
- Bieback, K., Kern, S., Kluter, H. & Eichler, H. Critical parameters for the isolation of mesenchymal stem cells from umbilical cord blood. *Stem Cells* **22**, 625–634 (2004).
- in 't Anker, P. S. et al. Isolation of mesenchymal stem cells of fetal or maternal origin from human placenta. *Stem Cells* **22**, 1338–1345 (2004).
- Zuk, P. A. et al. Multilineage cells from human adipose tissue: implications for cell-based therapies. *Tissue Eng.* **7**, 211–228 (2001).
- in 't Anker, P. S. et al. Amniotic fluid as a novel source of mesenchymal stem cells for therapeutic transplantation. *Blood* **102**, 1548–1549 (2003).
- Prockop, D. J. Marrow stromal cells as stem cells for nonhematopoietic tissues. *Science* **276**, 71–74 (1997).
- Horwitz, E. M. et al. Isolated allogeneic bone marrow-derived mesenchymal cells engraft and stimulate growth in children with osteogenesis imperfecta: implications for cell therapy of bone. *Proc. Natl Acad. Sci. USA* **99**, 8932–8937 (2002).
- Burt, R. K. et al. Clinical applications of blood-derived and marrow-derived stem cells for nonmalignant diseases. *J. Am. Med. Assoc.* **299**, 925–936 (2008).
- Di Nicola, M. et al. Human bone marrow stromal cells suppress T-lymphocyte proliferation induced by cellular or nonspecific mitogenic stimuli. *Blood* **99**, 3838–3843 (2002). This study showed the immunosuppressive properties of human bone-marrow stromal cells and became the catalyst for the use of such cells (now known as MSCs) to dampen inflammatory conditions, including graft-versus-host disease.
- Romieu-Mourez, R., Francois, M., Boivin, M. N., Stagg, J. & Galipeau, J. Regulation of MHC class II expression and antigen processing in murine and human mesenchymal stromal cells by IFN- $\gamma$ , TGF- $\beta$ , and cell density. *J. Immunol.* **179**, 1549–1558 (2007).
- Le Blanc, K., Tammik, C., Rosendahl, K., Zetterberg, E. & Ringden, O. HLA expression and immunologic properties of differentiated and undifferentiated mesenchymal stem cells. *Exp. Hematol.* **31**, 890–896 (2003).
- Miyara, M. & Sakaguchi, S. Natural regulatory T cells: mechanisms of suppression. *Trends Mol. Med.* **13**, 108–116 (2007).



37. Kawai, T. et al. HLA-mismatched renal transplantation without maintenance immunosuppression. *N. Engl. J. Med.* **358**, 353–361 (2008).  
This paper showed, in a clinical setting, proof of principle that tolerance can be induced to an allogeneic bone-marrow transplant, resulting in the complete withdrawal of immunosuppressive agents from the treatment regimens of these patients.
38. Hince, M. et al. The role of sex steroids and gonadectomy in the control of thymic involution. *Cell. Immunol.* doi:10.1016/j.cellimm.2007.10.007 (in the press).
39. Ildstad, S. T. & Sachs, D. H. Reconstitution with syngeneic plus allogeneic or xenogeneic bone marrow leads to specific acceptance of allografts or xenografts. *Nature* **307**, 168–170 (1984).
40. Ildstad, S. T., Wren, S. M., Bluestone, J. A., Barbieri, S. A. & Sachs, D. H. Characterization of mixed allogeneic chimeras. Immunocompetence, *in vitro* reactivity, and genetic specificity of tolerance. *J. Exp. Med.* **162**, 231–244 (1985).
41. Fandrich, F. et al. Preimplantation-stage stem cells induce long-term allogeneic graft acceptance without supplementary host conditioning. *Nature Med.* **8**, 171–178 (2002).
42. Kaufman, D. S., Hanson, E. T., Lewis, R. L., Auerbach, R. & Thomson, J. A. Hematopoietic colony-forming cells derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **98**, 10716–10721 (2001).
43. Burt, R. K. et al. Embryonic stem cells as an alternate marrow donor source: engraftment without graft-versus-host disease. *J. Exp. Med.* **199**, 895–904 (2004).
44. Verda, L. et al. Hematopoietic mixed chimerism derived from allogeneic embryonic stem cells prevents autoimmune diabetes mellitus in Nod mice. *Stem Cells* **26**, 381–386 (2008).
45. Alexander, S. I. et al. Chimerism and tolerance in a recipient of a deceased-donor liver transplant. *N. Engl. J. Med.* **358**, 369–374 (2008).
46. Scadden, D. T. The stem-cell niche as an entity of action. *Nature* **441**, 1075–1079 (2006).
47. Flores, K. G., Li, J., Sempowski, G. D., Haynes, B. F. & Hale, L. P. Analysis of the human thymic perivascular space during aging. *J. Clin. Invest.* **104**, 1031–1039 (1999).
48. Greenstein, B. D., Fitzpatrick, F. T., Kendall, M. D. & Wheeler, M. J. Regeneration of the thymus in old male rats treated with a stable analogue of LHRH. *J. Endocrinol.* **112**, 345–350 (1987).  
This paper provided, in an animal model, proof of principle that the thymus can be regenerated in an aged individual by using reversible chemical castration.
49. Olsen, N. J. & Kovacs, W. J. Effects of androgens on T and B lymphocyte development. *Immunol. Res.* **23**, 281–288 (2001).
50. Sutherland, J. S. et al. Activation of thymic regeneration in mice and humans following androgen blockade. *J. Immunol.* **175**, 2741–2753 (2005).  
This study showed a fundamentally new clinical approach for treating immunodeficiency states in humans, regenerating the immune system after bone-marrow transplantation or a damaging chemotherapy regimen.
51. Heng, T. S. et al. Effects of castration on thymocyte development in two different models of thymic involution. *J. Immunol.* **175**, 2982–2993 (2005).
52. Goldberg, G. L. et al. Enhanced immune reconstitution by sex steroid ablation following allogeneic hematopoietic stem cell transplantation. *J. Immunol.* **178**, 7473–7484 (2007).
53. Goldberg, G. L. et al. Sex steroid ablation enhances lymphoid recovery following autologous hematopoietic stem cell transplantation. *Transplantation* **80**, 1604–1613 (2005).
54. Sutherland, J. S. et al. Enhanced immune system regeneration in humans following allogeneic or autologous hematopoietic stem cell transplantation by temporary sex steroid blockade. *Clin. Cancer Res.* **14**, 1138–1149 (2008).
55. Min, D. et al. Sustained thymopoiesis and improvement in functional immunity induced by exogenous KGF administration in murine models of aging. *Blood* **109**, 2529–2537 (2007).
56. Min, D. et al. Protection from thymic epithelial cell injury by keratinocyte growth factor: a new approach to improve thymic and peripheral T-cell reconstitution after bone marrow transplantation. *Blood* **99**, 4592–4600 (2002).
57. Rossi, S. W. et al. Keratinocyte growth factor (KGF) enhances postnatal T-cell development via enhancements in proliferation and function of thymic epithelial cells. *Blood* **109**, 3803–3811 (2007).
58. Seggewiss, R. et al. Keratinocyte growth factor augments immune reconstitution after autologous hematopoietic progenitor cell transplantation in rhesus macaques. *Blood* **110**, 441–449 (2007).
59. Napolitano, L. A. et al. Growth hormone enhances thymic function in HIV-1-infected adults. *J. Clin. Invest.* **118**, 1085–1098 (2008).
60. Alpdogan, O. et al. Administration of interleukin-7 after allogeneic bone marrow transplantation improves immune reconstitution without aggravating graft-versus-host disease. *Blood* **98**, 2256–2265 (2001).
61. Wils, E. J. et al. Flt3 ligand expands lymphoid progenitors prior to recovery of thymopoiesis and accelerates T cell reconstitution after bone marrow transplantation. *J. Immunol.* **178**, 3551–3557 (2007).
62. Okamoto, Y., Douek, D. C., McFarland, R. D. & Koup, R. A. Effects of exogenous interleukin-7 on human thymus function. *Blood* **99**, 2851–2858 (2002).
63. Rutella, S., Danese, S. & Leone, G. Tolerogenic dendritic cells: cytokine modulation comes of age. *Blood* **108**, 1435–1440 (2006).
64. Jiang, X. X. et al. Human mesenchymal stem cells inhibit differentiation and function of monocyte-derived dendritic cells. *Blood* **105**, 4120–4126 (2005).
65. Zhang, W. et al. Effects of mesenchymal stem cells on differentiation, maturation, and function of human monocyte-derived dendritic cells. *Stem Cells Dev.* **13**, 263–271 (2004).
66. Sakaguchi, S. Naturally arising Foxp3-expressing CD25<sup>+</sup>CD4<sup>+</sup> regulatory T cells in immunological tolerance to self and non-self. *Nature Immunol.* **6**, 345–352 (2005).
67. Roncarolo, M. G. et al. Interleukin-10-secreting type 1 regulatory T cells in rodents and humans. *Immunol. Rev.* **212**, 28–50 (2006).
68. Joffre, O. et al. Prevention of acute and chronic allograft rejection with CD4<sup>+</sup>CD25<sup>+</sup>Foxp3<sup>+</sup> regulatory T lymphocytes. *Nature Med.* **14**, 88–92 (2008).
69. Yamazaki, S., Inaba, K., Tarbell, K. V. & Steinman, R. M. Dendritic cells expand antigen-specific Foxp3<sup>+</sup>CD25<sup>+</sup>CD4<sup>+</sup> regulatory T cells including suppressors of alloreactivity. *Immunol. Rev.* **212**, 314–329 (2006).
70. Yamazaki, S. et al. Effective expansion of alloantigen-specific Foxp3<sup>+</sup>CD25<sup>+</sup>CD4<sup>+</sup> regulatory T cells by dendritic cells during the mixed leukocyte reaction. *Proc. Natl Acad. Sci. USA* **103**, 2758–2763 (2006).
71. Zhan, X. et al. Functional antigen-presenting leucocytes derived from human embryonic stem cells *in vitro*. *Lancet* **364**, 163–171 (2004).
72. Lee, M. K. et al. Promotion of allograft survival by CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells: evidence for *in vivo* inhibition of effector cell proliferation. *J. Immunol.* **172**, 6539–6544 (2004).
73. Byrne, J. A. et al. Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* **450**, 497–502 (2007).
74. French, A. J. et al. Development of human cloned blastocysts following somatic cell nuclear transfer with adult fibroblasts. *Stem Cells* **26**, 485–493 (2008).
75. French, A. J., Wood, S. H. & Trounson, A. O. Human therapeutic cloning (NTSC): applying research from mammalian reproductive cloning. *Stem Cell Rev.* **2**, 265–276 (2006).
76. Park, I. H. et al. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
77. Takahashi, K., Okita, K., Nakagawa, M. & Yamanaka, S. Induction of pluripotent stem cells from fibroblast cultures. *Nature Protocols* **2**, 3081–3089 (2007).
78. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).  
This study induced pluripotency in somatic cells from adult humans, enabling patient-specific and disease-specific iPS cells to be generated and raising the possibility of their use in regenerative medicine.
79. Nakagawa, M. et al. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26**, 101–106 (2008).  
This paper described an important step in adapting iPS cells for clinical application by removing the requirement for transfection with tumour-inducing genes such as Myc.
80. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
81. Anderson, G., Lane, P. J. & Jenkinson, E. J. Generating intrathymic microenvironments to establish T-cell tolerance. *Nature Rev. Immunol.* **7**, 954–963 (2007).
82. Derbinski, J., Schulte, A., Kyewski, B. & Klein, L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nature Immunol.* **2**, 1032–1039 (2001).
83. Gabler, J., Arnold, J. & Kyewski, B. Promiscuous gene expression and the developmental dynamics of medullary thymic epithelial cells. *Eur. J. Immunol.* **37**, 3363–3372 (2007).
84. Gillard, G. O. & Farr, A. G. Features of medullary thymic epithelium implicate postnatal development in maintaining epithelial heterogeneity and tissue-restricted antigen expression. *J. Immunol.* **176**, 5815–5824 (2006).
85. Rossi, S. W. et al. RANK signals from CD4<sup>+</sup>3<sup>+</sup> inducer cells regulate development of Aire-expressing epithelial cells in the thymic medulla. *J. Exp. Med.* **204**, 1267–1272 (2007).
86. Sakaguchi, S. et al. Foxp3<sup>+</sup>CD25<sup>+</sup>CD4<sup>+</sup> natural regulatory T cells in dominant self-tolerance and autoimmune disease. *Immunol. Rev.* **212**, 8–27 (2006).
87. Spence, P. J. & Green, E. A. Foxp3<sup>+</sup> regulatory T cells promiscuously accept thymic signals critical for their development. *Proc. Natl Acad. Sci. USA* **105**, 973–978 (2008).

**Acknowledgements** We acknowledge generous research support from the National Health and Medical Research Council of Australia, the Australian Stem Cell Centre and Norwood Immunology.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to A.C. ([ann.chidgey@med.monash.edu.au](mailto:ann.chidgey@med.monash.edu.au)).



# A chemical approach to stem-cell biology and regenerative medicine

Yue Xu<sup>1</sup>, Yan Shi<sup>1</sup> & Sheng Ding<sup>1</sup>

**An improved understanding of stem-cell and regenerative biology, as well as a better control of stem-cell fate, is likely to produce treatments for many devastating diseases and injuries. Chemical approaches are starting to have an increasingly important role in this young field. Attention has focused on chemical approaches that allow the precise manipulation of cells *in vitro* to obtain homogeneous cell types for cell-based therapies. Another promising approach is the development of conventional chemical and biological therapeutics to stimulate endogenous cells to regenerate. Such therapeutics can act on target cells or their niches *in vivo* to promote cell survival, proliferation, differentiation, reprogramming and homing.**

Recent breakthroughs in stem-cell biology have generated enthusiasm for therapeutic developments towards tissue and organ repair and/or regeneration. These advances include insights into the intrinsic mechanisms<sup>1–4</sup> and the niche interactions<sup>5</sup> that regulate the fate of embryonic stem (ES) cells and adult stem cells. Other recent advances are the ability to manipulate and control stem-cell fate and function in a more precise and defined manner<sup>6–9</sup> and the ability to reprogramme lineage-committed cells to revert to more-primitive multipotent<sup>10</sup> states (in which the cells can form many cell types) or even pluripotent states<sup>11</sup> (in which the cells can form all the cell types of the body). Not only are various types of stem cell excellent model systems for studying the fundamental biology of human development and diseases, but they also provide superior vehicles compared with conventional cell lines for drug discovery (from disease modelling, target validation and high-throughput screening to development of clinical candidates). Furthermore, stem cells can be used to replace cells that have been damaged as a result of disease and injury. They can also be endogenous targets for conventional small-molecule or biological therapeutics that stimulate a body's own cells to regenerate *in vivo*.

Chemical approaches (Fig. 1) and small molecules have provided the key to many biological discoveries. Understanding of a biological phenomenon often begins by observing it and then discovering or designing ways to perturb it. Although genetic methods have been widely used for this purpose, a chemical approach offers distinct advantages. For example, small molecules typically provide a high degree of temporal control over protein function, inducing rapid inhibition or activation, and their effects are often reversible. Their effects can also be finely tuned by varying the concentrations of the small molecule. Furthermore, a single small molecule can simultaneously modulate multiple specific targets within a protein family or across different protein families. This allows the production of a desirable phenotype in a synergistically favourable manner. It also has considerable advantages when viewed from a primary screening and clinical-development perspective.

Examples of small molecules that modulate single or multiple targets<sup>12</sup> are retinoic acid, cytidine analogues, histone-deacetylase inhibitors and protein kinase inhibitors (Fig. 2). These chemicals have been used to modulate and dissect stem-cell phenotypes and may well have a role in treatment of stem- and progenitor-cell-based diseases such as some types of cancer.

In this review, we discuss chemical approaches to stem-cell and regenerative biology, including the identification of new chemical entities or chemically defined conditions for studying cell-fate regulation *in vitro* and *in vivo*. We also discuss ways in which such chemical approaches enhance cell-replacement therapy and the use of therapeutic molecules to control endogenous cell fate *in vivo* for the treatment of degenerative diseases, injuries, ageing and cancers.

## Stem-cell modulation *in vitro*

The ability to isolate stem cells and progenitor cells from tissues and manipulate them *in vitro* gives rise to several attractive opportunities. It can provide model systems for understanding the intrinsic regulation of these cells and their interactions with exogenous signals, including those underlying development and disease. Moreover, controlling self-renewal expansion and differentiation will be necessary to generate sufficient homogeneous functional cell types for cell-based therapies. And such homogeneous and physiologically relevant cell populations would also be useful for drug discovery.

## Self-renewal

Pluripotent stem cells give rise to all cell types in a body. These typically include the following: conventional ES cells derived from the inner cell mass of pre-implantation embryos<sup>13</sup>; germline stem cells and derivatives<sup>14–16</sup>; epiblast stem cells derived from post-implantation epiblast-stage embryos<sup>17,18</sup>; and induced pluripotent stem (iPS) cells derived from somatic cells<sup>11</sup>. Because ES cells are an excellent model for pluripotent stem cells, we focus on the regulation of ES cells by chemical approaches.

Conventionally, ES cells are maintained and expanded in the presence of feeder cells, serum and additional exogenous factors<sup>13</sup> such as leukaemia inhibitory factor (LIF) for mouse ES cells and basic fibroblast growth factor (bFGF; also known as FGF2) for human ES cells. The use of such conventional culture conditions has presented a number of problems. First, such conditions are highly variable both in the way feeder cells are used and in the composition of serum products. Thus, maintaining large-scale, consistent and robust long-term ES-cell cultures has been a challenge. Second, unknown factors from feeder cells and/or serum, as well as certain defined exogenous factors, may bias ES cells towards having specific lineage differentiation properties. They

<sup>1</sup>Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

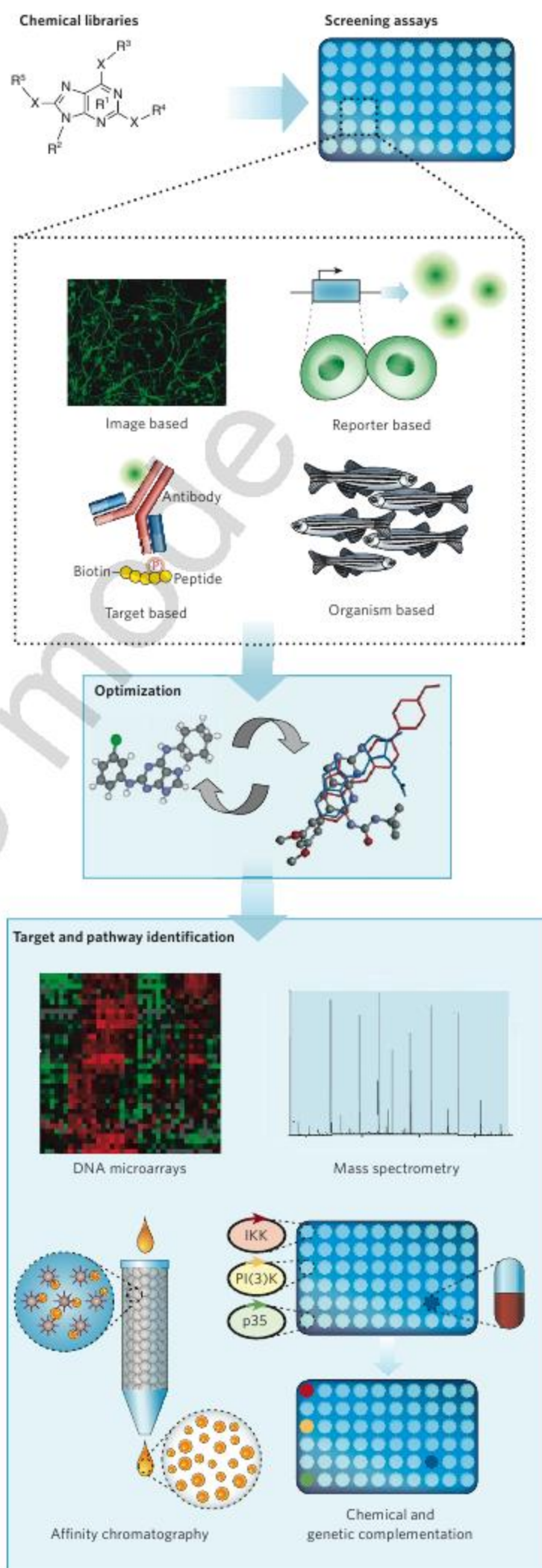


do this by modulating additional gene expression that coexists with the basic pluripotency network. This adds to the existing complexity of ES cell lines, especially in humans. Different ES cell lines show different self-renewal properties, as well as showing different propensities to differentiate into different lineages (or different subtypes of the same lineage) owing to their own unique genetic and epigenetic background<sup>19</sup>. Third, unknown factors may operate in conjunction with a specific treatment to change the outcome of a particular cellular process.

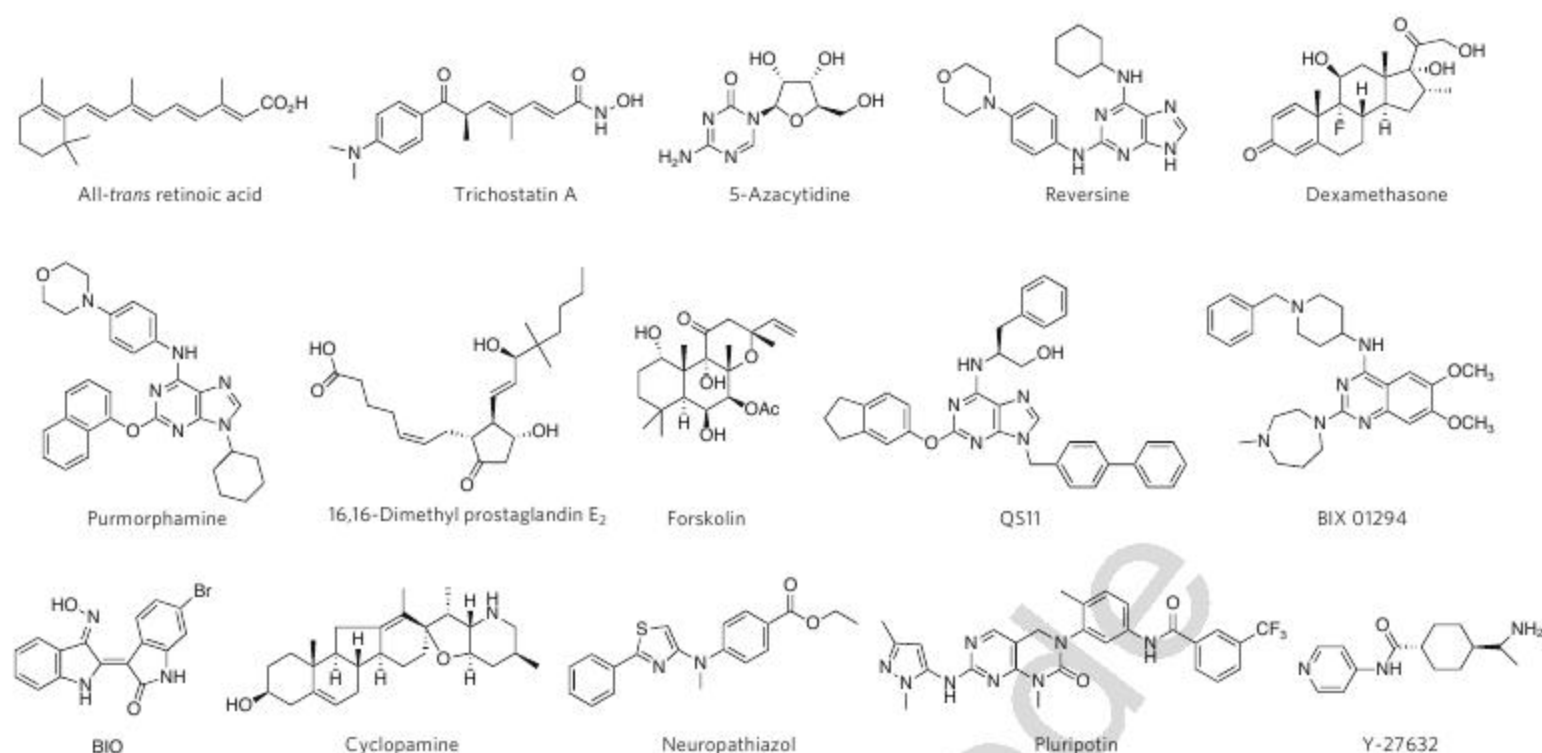
To address these problems, we carried out a high-throughput screen of 50,000 synthetic small molecules using a transgenic reporter mouse ES cell line expressing green fluorescent protein (GFP) under control of the *Oct4* gene (a gene specifically expressed in pluripotent stem cells) regulatory elements. The screen was conducted in the absence of feeder cells, serum and LIF, to search for small molecules that can maintain self-renewal of ES cells in chemically defined conditions. OCT4–GFP expression and the characteristic compact domed colony morphology of mouse ES cells were used as criteria to select primary hits in the screen. Secondary confirmation assays and structure–activity–relationship studies led to the identification of a novel compound named pluripotin (also known as SC1)<sup>6</sup>. Pluripotin sustains homogeneous self-renewal of mouse ES cells long-term in chemically defined medium conditions in the absence of feeder cells, serum, LIF and bone morphogenetic proteins (BMPs). The cells remain pluripotent in serially passaged culture *in vitro* without losing their germline transmission ability *in vivo*<sup>6</sup>. Furthermore, pluripotin functions independently of exogenous activation of LIF–STAT3, BMP–SMAD-ID and WNT– $\beta$ -catenin pathways. Through affinity pull-down experiments using a pluripotin-immobilized matrix, the molecular targets of pluripotin were identified as RasGAP and extracellular-signal-regulated kinase 1 (ERK1), two endogenously expressed proteins with differentiation-inducing activity. Additional biochemical, genetic and pharmacological studies have shown that simultaneous inhibition of ERK1 and RasGAP by pluripotin or other independent methods is sufficient for long-term self-renewal of mouse ES cells<sup>6</sup>.

From a small-molecule and drug-discovery perspective, this study demonstrated that controlling a complex phenotype (or curing a disease) can require modulation of more than one target. A single small molecule with the desirable, specific polypharmacological activity can be selected out through a rationally designed phenotypic screen. From

**Figure 1 | A screening approach.** Chemical libraries for biological screens can be assembled from synthetic compounds and natural products through chemical synthesis and/or from commercial sources on the basis of their biological fitness and molecular diversity. Screening against a defined molecular target usually exploits a protein's function (for example, enzymatic activity or signal transduction) or molecular interactions with its partners. Phenotypic screens are powerful, in that prior knowledge of defined targets and mechanisms for the given desirable phenotype is not required. Such screens can be carried out in cells or whole organisms by examining multiple markers and functional changes (for example, cell morphology and behaviour) using automated high-content imaging technologies in a high-throughput manner (see page 345). After hit compounds are identified from a primary screen, they are typically confirmed by using a series of secondary assays, which are more functional, as well as being analysed by informatics tools. Before a compound enters mechanistic studies, it is normally optimized through structure–activity–relationship studies to improve its properties, such as increasing its potency, specificity and enhancing its pharmacokinetic properties. To identify the molecular targets and pathways of an unknown compound, affinity chromatography with labelled compounds in conjunction with mass spectrometry is most commonly used. Target proteins are typically validated by gain- and loss-of-function studies (for example, through microarray analysis and RNA interference) to recapitulate the compound's activity, as well as by using other relevant biochemical and cellular assays (shown here as chemical and genetic complementation). Unbiased expression analysis and genetic epistasis may also be used to provide mechanistic insights. R<sup>1</sup> denotes a heterocycle; X denotes O, N, S or C. IKK, inhibitor of nuclear factor- $\kappa$ B (I $\kappa$ B) kinase; PI(3)K, phosphatidylinositol-3-OH kinase.







**Figure 2 | Selected chemical compounds that regulate cell fate.** Shown are synthetic small molecules and natural products that bind to nuclear receptors (all-*trans* retinoic acid and dexamethasone), histone-modifying enzymes and DNA-modifying enzymes (trichostatin A, BIX 01294

and 5-azacytidine), and protein kinases and signalling molecules (reversine, purmorphamine, 16,16-dimethyl prostaglandin E<sub>2</sub>, forskolin, QS11, BIO, cyclopamine, neuropathiazol, pluripotin and Y-27632).

a stem-cell perspective, it suggests that self-renewal of ES cells, and perhaps other types of stem cell as well, may be largely driven by intrinsic regulators in the cell and does not require activation of additional pathways by exogenous factors, such as LIF or BMPs<sup>20</sup>.

Maintenance of self-renewal in stem cells can be simply viewed as involving multiple processes of continued proliferation, as well as inhibiting differentiation and cell death. This requires a fine balance and cross-regulation between positive and negative regulators. Stem cells might autonomously express almost all gene products that are essential for self-renewal. However, endogenous expression of some differentiation-inducing genes at a certain level in the undifferentiated ES cells could cause differentiation in culture conditions that do not inhibit their negative effects. Therefore, the key to achieving self-renewal in ES cells may be to inhibit the negative effects of endogenously expressed, pluripotency inhibitory proteins (for example, proteins involved in differentiation or cell death).

A balanced self-renewal state can also be achieved through a specific combination of multiple pathway activation by exogenous factors (for example, LIF, BMPs and WNT proteins for mouse ES cells<sup>20</sup>, or bFGF, activin and WNT proteins for human ES cells<sup>21,22</sup>). Although these exogenous factors at the appropriate concentrations inhibit each other's differentiation activity, they can also mediate lineage-specific gene expression, which coexists with the pluripotency gene networks, or induce specific lineage differentiation. This idea is also supported by an independent study in which a combination of three specific chemical inhibitors of the protein kinases FGF receptor (FGFR), mitogen-activated protein kinase kinase (MEK) and glycogen-synthase kinase 3 (GSK3) supports derivation and long-term self-renewal of mouse ES cells in the absence of exogenous cytokines<sup>68</sup>. Not only are these chemicals useful for defining cell-culture conditions and regulatory mechanisms (replacing exogenous factors and providing a platform for supporting a more reliable and robust cell culture), but they also present an opportunity for derivation of new cell lines from strains or species for which this has been difficult. For example, a combination of the three inhibitors of FGFR, MEK and GSK3 was used to derive ES cells from the mouse strain CBA and *Stat3*<sup>-/-</sup> mice.

Small molecules can function intracellularly to inhibit key differentiation-inducing proteins, thereby bypassing the need to express receptors

for exogenous factors that inhibit differentiation. Positive results have been obtained using chemical approaches to derive pluripotent stem cells from other species (for example, rat and human) that resemble more closely the conventional inner-cell-mass-stage mouse ES cells. This is also in line with the recent finding<sup>17,18</sup> that the conventional human ES cells represent a later epiblast stage of pluripotent cells and have defining features in common with rodent epiblast stem cells.

Various exogenous growth factors, including bFGF<sup>7,22</sup>, TGF- $\beta$  and activin<sup>21,23</sup>, WNT proteins<sup>7,24</sup> and insulin-like growth factor<sup>25</sup>, support self-renewal of human ES cells in chemically defined conditions. Identification of small molecules that function similarly to pluripotin in mouse ES cells may lead to improved control of human ES-cell fate. For example, the extremely poor survival of dissociated single human ES cells had been a problem for routine large-scale culture and clonal selection of human ES cells<sup>26</sup>. Testing of caspase and protein-kinase inhibitors identified an inhibitor of the protein kinase ROCK, known as Y-27632, which allows survival of dissociated human ES cells. Recently, we carried out high-throughput screens of large-scale combinatorial chemical libraries in human ES cells and identified distinct small molecules that can replace bFGF in maintaining long-term self-renewal of human ES cells in a feeder-free and serum-free chemically defined condition or can potentially promote single-cell survival of human ES cells. Further characterization of the mechanisms involved in the action of these small molecules will provide a better understanding of human ES-cell biology.

In contrast with the progress made with ES cells, maintaining long-term self-renewal of multipotent tissue-specific stem cells (Fig. 3Aa), especially at the most primitive stage, remains a challenge. For example, long-term haematopoietic stem cells (HSCs) quickly lose their ability to serially repopulate the entire haematopoietic system during *in vitro* culture, despite the use of multiple exogenous protein factors. Similarly, conditions with growth factors (for example, bFGF and/or EGF) can only expand primitive neural stem cells for a limited number of generations *in vitro*. In such conditions, they typically become more glia-restricted and lose the ability to be patterned to subtype-specific neuronal types<sup>27</sup>. Finding small molecules that can inhibit their differentiation would be useful.



To this end, a recent study<sup>28</sup> used such a chemical approach to study multipotent islet1 (Isl1)<sup>+</sup> cardiovascular progenitor cells (MICPs), which reside in both embryonic and adult hearts. MICPs can generate the three main heart cell types: cardiac muscle, smooth muscle and endothelial cells<sup>29</sup>. A high-throughput screen was carried out to identify small molecules capable of inducing expansion of MICPs; Isl1-LacZ was used as a marker. Several compounds were identified that significantly increased MICP expansion, with the potent GSK3 inhibitor BIO being one of the strongest. This discovery led to a series of *in vitro* and *in vivo* studies of the role of the WNT- $\beta$ -catenin signal (as secreted from the cardiac mesenchymal niche cells), which functions in a sequential manner first to block the specification of the MICPs from mesodermal precursor cells and then to promote their self-renewal by inhibiting their further differentiation into cardiomyocyte and smooth muscle cells. Interestingly, BIO treatment also induced expansion of human Isl1<sup>+</sup> cardiovascular progenitor cells, suggesting a conserved role of WNT-mediated signalling in self-renewal of MICPs.

### Differentiation

Conventional differentiation of ES cells has been an inefficient and nonspecific process typically involving co-culture with feeder cells<sup>30,31</sup> or growing them in suspension to form embryoid bodies in the presence of complex serum and additional factors. The cell type of interest is then selected from a heterogeneous cell population by the expression of markers. The development of small molecules and/or chemically defined conditions for more effective stem-cell differentiation has attracted significant efforts, for practical reasons and for improving the understanding of the mechanisms regulating differentiation.

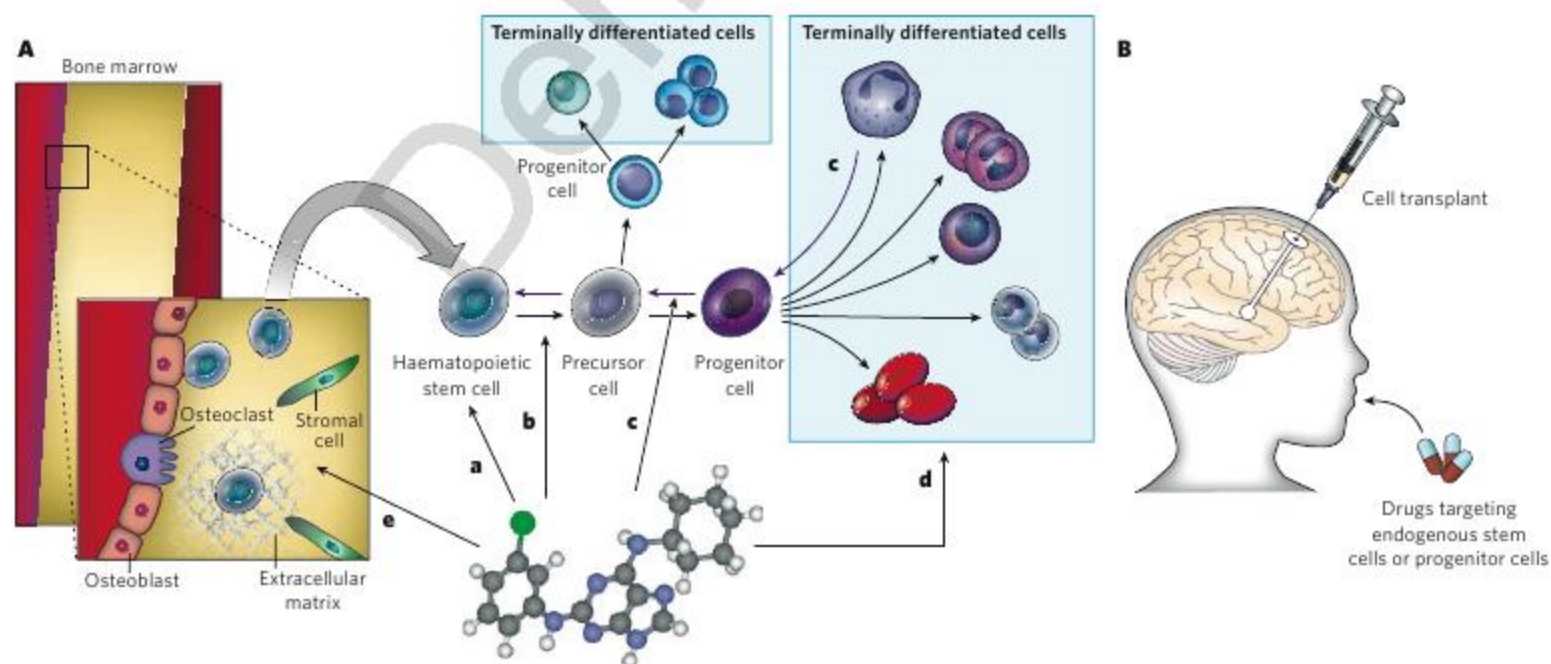
Advances in developmental biology have guided the design of directed, stepwise differentiation of stem cells in ways that recapitulate the progression of embryonic development. This was achieved by sequentially treating cells with combinations of embryonic signalling molecules. An early elegant example was the generation of functional motor neurons from mouse ES cells by first neuralizing and caudalizing (differentiating and patterning) cells with retinoic acid. Subsequently, the cells were ventralized (positioned) by a specific small-molecule agonist of the hedgehog-mediated signalling pathway<sup>32</sup>. Conceptually similar, but technically more sophisticated, methods have since been developed for directing human ES cells to differentiate along the

following paths: neural precursor cells  $\rightarrow$  subtype-specific neuronal progenitor cells  $\rightarrow$  mature functional neurons<sup>33,34</sup>; mesendoderm  $\rightarrow$  mesoderm  $\rightarrow$  cardiovascular precursor cells  $\rightarrow$  immature cardiac cell types  $\rightarrow$  mature cardiomyocytes<sup>35,36</sup>; and mesendoderm  $\rightarrow$  definitive endoderm  $\rightarrow$  primitive gut tube  $\rightarrow$  posterior foregut  $\rightarrow$  pancreatic endoderm  $\rightarrow$  endocrine precursor cells  $\rightarrow$  hormone-producing endocrine cells<sup>9,37</sup>.

To substantially increase the homogeneity, functionality and yield of the intermediate progenitor cells or terminally differentiated cell types, a more precise knowledge of lineage specification will be needed. Thus it would be highly desirable to identify additional small molecules (Fig. 3Ab) that can function synergistically with factors that have already been defined, such as FGF and WNT proteins. The ultimate goal is the development of specific, efficient and completely chemically defined medium conditions. Towards this goal, we have used high-content imaging-based analysis of immunostained cells with a neuron-specific marker to identify a novel synthetic small molecule named neuropathiazol from a large chemical library. This molecule can specifically induce robust differentiation of primary multipotent hippocampal neural progenitor cells into neurons, even in gliogenic conditions<sup>38</sup>. More recently, screening of libraries of pharmacologically active compounds using neurospheres and simple homogeneous assays has identified drugs and endogenous metabolites for inhibiting neurosphere proliferation<sup>39</sup>, and/or promoting neuronal differentiation and/or survival<sup>40</sup>. Some of those small molecules — as well as others such as histone-deacetylase inhibitors<sup>41</sup> and serotonin-uptake blockers<sup>42</sup> — promote neurogenesis in cell culture, as well as *in vivo*, with specific behaviour modulation when administered directly. These studies have demonstrated the great utility of chemical approaches in regulating cell fate for *in vitro* applications. Even more importantly, they hold promise that relevant cell-fate modulation *in vitro* by small molecules will provide models and leads for therapeutic development towards *in vivo* regeneration.

### Reprogramming

Mammalian tissue-specific stem cells and progenitor cells typically have restricted developmental potential *in vivo*. Such cells can only differentiate towards cell types with a more restricted potential within the same lineage boundary. However, in certain conditions, mammalian



**Figure 3 | Therapeutic strategies for regenerative medicine.** A, Small molecules can target stem cells or progenitor cells for self-renewal (a) or differentiation (b). Small molecules can also target lineage-restricted cells for the generation of more-primitive cells or other tissue cell types (c), or they can regulate the survival, proliferation, homing or reprogramming of terminally differentiated cells (d). They can also control the cell fate

and/or function of target cells by regulating the appropriate cellular niches (e). B, Homogeneous and functional cell types generated in chemically defined conditions can be used for cell-based therapy; and alternatively, conventional chemical and biological therapeutics can be developed to target patients' own cells or their niches to stimulate regeneration *in vivo*.



cells can be reprogrammed to adopt alternative cell fates across lineage boundaries or to return to more-primitive states *in vitro*, as well as *in vivo*, through various mechanisms. Such techniques include somatic cell nuclear transfer (to generate totipotent cells, which are sufficient to form an entire organism)<sup>43,44</sup>, cell (or cell extract) fusion<sup>45–47</sup>, genetic alterations<sup>11,48,49</sup> or defined exogenous molecules<sup>10,50,51</sup>. At the molecular level, these studies suggest that a stably balanced epigenetic state can be shifted by specific treatments. Recent successes in generating ES cells by somatic cell nuclear transfer into mitotic zygotes from mice<sup>52</sup> and into oocytes from non-human primates<sup>53</sup> suggest that it might be possible to generate human ES cells by using a similar approach, which might lessen ethical concerns.

Another breakthrough in reprogramming is the identification of an easy genetic approach to convert somatic cells back to iPS cells<sup>11,54–59</sup> (see page 322). The simplicity of this genetic approach has opened up tremendous opportunities to generate patient-specific cells for various applications (for example, cell-based therapy or drug discovery) without the controversies associated with the conventional human ES cells. In addition, it facilitates studies of the intriguing epigenetic reversal process on a fully defined basis.

Practical clinical applications of the iPS-cell approach will largely depend on the resolution of two crucial issues. The first is elimination of the risks associated with exogenous genetic manipulations, as well as possible endogenous genetic alterations, during the slow and inefficient reprogramming process. The second is the generation of homogeneous populations of lineage-specific cell types from iPS cells. One strategy is to replace the genetic method with chemically defined approaches, such as protein or RNA transduction and/or growth factor or small-molecule treatment. We have explored two approaches for replacing viral transduction of transcription factors. The first approach was to test methods that induce pluripotency in different cell types on the basis of the idea that certain accessible cell types may endogenously express or have less silenced loci of some of the required genes for inducing pluripotency. Thus the cells can be more efficiently reprogrammed with fewer genetic manipulations. The second approach was to identify defined exogenous factors and small molecules that can activate a pluripotency network, inhibit negative regulators of pluripotency or directly regulate chromatin modifications. Both approaches independently or in conjunction have generated positive results that may ultimately lead to development of fully chemically defined conditions for the generation of iPS cells.

Given the difficulties with differentiating pluripotent stem cells for specific applications, an attractive approach to reprogramming is to generate intermediate lineage-specific stem cells or progenitor cells or other types of differentiated functional cell through lineage reversal or transdifferentiation (Fig. 3Ac). Although problems with genetic manipulations (if used), and efficiency and heterogeneity, would still need to be resolved, these reprogrammed cells might be advantageous for differentiation and/or have a lower cancer risk because they are lineage restricted and do not form teratomas *in vivo*.

Reprogramming of somatic cells to express previously silenced genes or to become another cell type has been extensively studied, providing the conceptual and technical basis for epigenetic reprogramming biology. Examples of such *in vitro* studies include heterokaryon (a fusion of two different cell types) formation-induced reprogramming, which is independent of cell division<sup>60</sup>, phenotypic conversion of certain pancreatic cells to hepatocytes<sup>61</sup> by dexamethasone treatment or by overexpression of the transcription factor C/EBP- $\beta$ , conversion of hepatocytes to a pancreatic phenotype by overexpression<sup>61</sup> of PDX1, reprogramming of B cells through the transcription factor Pax5 deletion<sup>48,49</sup>, reprogramming of normally glia-restricted oligodendrocyte precursor cells to multipotent neural precursor cells by sequential treatment with BMPs and bFGF<sup>50</sup>, and mesenchymal-cell reprogramming by controlling matrix elasticity<sup>62</sup>.

To screen rationally for small molecules that would reprogramme lineage-committed cells to become more-primitive precursor cells, we carried out a cell-based functional screen based on the idea that lineage-reversed cells would regain multipotency. Lineage-committed myoblasts

were screened using a two-stage screening protocol, in which cells were initially treated with libraries of compounds to induce dedifferentiation. They were then assayed for their ability to differentiate into otherwise non-permitted osteoblasts in osteogenic conditions. A synthetic small molecule named reversine was identified and shown to reprogramme multiple lineage-restricted cell types to a more-primitive multipotent state at the clonal level<sup>10,63</sup>. The cellular targets of reversine were identified through affinity chromatography as MEK1 and non-muscle myosin II heavy chain. Mechanistic studies suggested that inhibition of both target proteins' activities is required, which entails cell-cycle G2–M phase staging, cytoskeletal reorganization and cell signalling modulation.

### Targets for therapeutic interventions

Cell-based therapy using tissue-specific cells either isolated from donors or derived from pluripotent stem cells holds promise for treatment of many devastating diseases and injuries. Clinical successes have included several cell-replacement therapies, such as HSC transplantation for blood-related disorders and pancreatic islet or  $\beta$ -cell transplantation for type 1 diabetes. However, challenging issues remain. It is difficult to obtain donor cells, and there are issues with immunological compatibility and the precise control of cell fate in defined conditions *ex vivo*. As highlighted above, chemical approaches would support cell-based therapy by generating and/or enhancing transplantable cells through better *ex vivo* control of cell survival, growth, differentiation and reprogramming.

Given that many adult tissues and organs have endogenous stem cells and progenitor cells that participate in normal tissue homeostasis and regenerative processes in response to injuries, it is conceivable that the body's own cells can be targeted *in vivo* to enhance regeneration. As the scientific understanding of adult stem-cell biology increases, this approach will be further strengthened by the decades of experience in development of conventional drugs. Consequently, endogenous stem cells and progenitor cells and their cellular niches are targets for therapeutic development.

To identify new mechanisms and small molecules that can influence endogenous stem-cell behaviour, phenotypic functional screens at the cellular or organismal level have been particularly useful. A recent phenotypic screen of known drug collections in zebrafish embryos identified small-molecule regulators of prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) synthesis that modulate HSC numbers *in vivo*<sup>64</sup>. This discovery was further extended to show that a stabilized PGE<sub>2</sub> analogue, 16,16-dimethyl PGE<sub>2</sub>, improved kidney-marrow recovery after irradiation injury in the adult zebrafish. *Ex vivo* treatment of mouse bone-marrow cells with 16,16-dimethyl PGE<sub>2</sub> increased the frequency of long-term HSCs present in mouse bone marrow after limiting-dilution competitive transplantation<sup>64</sup>. This suggests that PGE<sub>2</sub> functions as a potent regulator of vertebrate HSC homeostasis, and its pathway modulation by small molecules may be useful for treating patients undergoing bone-marrow transplantation.

Another strategy to develop small molecules for regeneration is to focus on defined molecular targets or pathways, such as WNT- and hedgehog-mediated signalling, which are implicated in specific regenerative processes. One potential concern about activating a regenerative pathway is the risk of causing cancer, as certain genetic alterations or abnormal gene expressions that lead to the activation of some of those pathways have been associated with cancer. Strategies such as temporal and/or synergistic activation may provide a viable solution. To identify novel compounds and pathways that interact with the canonical WNT- $\beta$ -catenin signalling pathway, we recently carried out a reporter-based high-throughput screen for molecules that synergistically activate the reporter in the presence of WNT3A. A 2,6,9-trisubstituted purine compound, QS11, was found to synergize with canonical WNT proteins both *in vitro* and *in vivo*<sup>65</sup>. Affinity chromatography identified ARF-GAP as a target of QS11. Additional biochemical, genetic and functional studies have established that QS11 inhibits ARF-GAP, and as a consequence modulates ARF activity and  $\beta$ -catenin localization so that it crosstalks with WNT-mediated signalling.



Significant advances in defining adult stem-cell niches and understanding how they regulate stem-cell function *in vivo* have provided new strategies for controlling cell fate by pharmacologically manipulating the niches<sup>5</sup> (Fig. 3Ae). Such approaches alone or in combination with approaches to target stem cells directly may provide broader selection of molecular targets or elicit more robust response in specific settings. A proof-of-concept demonstration in mice showed that daily treatment with a parathyroid hormone that augments osteoblasts, a key component of the HSC niche, resulted in expansion and protection of HSCs with therapeutic benefit in three clinically relevant models of stem-cell-based therapy<sup>66</sup>. In another example, regeneration of damaged axons in the adult central nervous system is restricted by both their intrinsic incompetence and microenvironment, consisting of myelin and the glial scar. A small-molecule screen for compounds that can neutralize inhibitory activity, such as that associated with central-nervous-system myelin, identified EGFR inhibitors as potent promoters of neurite outgrowth on cerebellar granule neurons on an immobilized myelin substrate<sup>67</sup>. More importantly, local administration of EGFR inhibitors led to significant regeneration of injured optic nerve fibres in mice, suggesting that targeting the inhibitory microenvironment may provide a therapeutic avenue for enhancing axon regeneration after central-nervous-system injury.

## Perspectives

Stem cells present enormous opportunities for basic research, drug discovery and therapies. Conventional small-molecule or biological therapeutics will probably become a more convenient form of regenerative medicine, working to unleash the body's own regenerative capacities by promoting survival, migration, proliferation, differentiation or reprogramming of endogenous cells.

Continued development and application of chemical approaches in stem cells will undoubtedly lead to identification of additional small molecules and more precisely defined and 'individualized' conditions for controlling cell fate *in vitro* and *in vivo*.

- Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).  
This paper reports a genome-wide location analysis of the genes *Oct4*, *Sox2* and *Nanog*, and proposes a model of core ES-cell regulatory circuitry for maintaining the pluripotent state of ES cells.
- Lee, T. I. *et al.* Control of developmental regulator's by polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
- Scadden, D. T. The stem-cell niche as an entity of action. *Nature* **441**, 1075–1079 (2006).
- Chen, S. B. *et al.* Self-renewal of embryonic stem cells by a small molecule. *Proc. Natl Acad. Sci. USA* **103**, 17266–17271 (2006).  
This paper describes the identification of a novel synthetic small molecule that can maintain long-term self-renewal of mouse ES cells in the absence of feeder cells, serum, LIF, BMPs and WNT proteins.
- Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nature Biotechnol.* **24**, 185–187 (2006).
- Yao, S. *et al.* Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. *Proc. Natl Acad. Sci. USA* **103**, 6907–6912 (2006).
- D'Amour, K. A. *et al.* Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature Biotechnol.* **24**, 1392–1401 (2006).  
This paper elegantly describes a directed, stepwise differentiation of human ES cells into functional pancreatic hormone-expressing endocrine cells.
- Chen, S. B., Zhang, Q. S., Wu, X., Schultz, P. G. & Ding, S. Dedifferentiation of lineage-committed cells by a small molecule. *J. Am. Chem. Soc.* **126**, 410–411 (2004).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).  
This paper was the first to demonstrate that mouse somatic cells can be reprogrammed to become iPS cells by viral transduction of four defined factors: OCT4, SOX2, KLF4 and Myc.
- Ding, S. & Schultz, P. G. A role for chemistry in stem cell biology. *Nature Biotechnol.* **22**, 833–840 (2004).
- Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev.* **19**, 1129–1155 (2005).
- Shamblott, M. J. *et al.* Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl Acad. Sci. USA* **95**, 13726–13731 (1998).
- Kanatsu-Shinohara, M. *et al.* Generation of pluripotent stem cells from neonatal mouse testis. *Cell* **119**, 1001–1012 (2004).
- Guan, K. *et al.* Pluripotency of spermatogonial stem cells from adult mouse testis. *Nature* **440**, 1199–1203 (2006).
- Brons, J. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
- Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).  
References 17 and 18 report the derivation of pluripotent epiblast stem cells from post-implantation, epiblast-stage embryos of mice and rats.
- Wu, H. *et al.* Integrative genomic and functional analyses reveal neuronal subtype differentiation bias in human embryonic stem cell lines. *Proc. Natl Acad. Sci. USA* **104**, 13821–13826 (2007).
- Ying, Q. L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281–292 (2003).
- Vallier, L., Reynolds, D. & Pederson, R. A. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev. Biol.* **275**, 403–421 (2004).
- Xu, R. H. *et al.* Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature Methods* **2**, 185–190 (2005).
- Beattie, G. M. *et al.* Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem Cells* **23**, 489–495 (2005).
- Lu, J., Hou, R., Booth, C. J., Yang, S.-H. & Snyder, M. Defined culture conditions of human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **103**, 5688–5693 (2006).
- Wang, L. *et al.* Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* **110**, 4111–4119 (2007).
- Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature Biotechnol.* **25**, 681–686 (2007).
- Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, 1594–1606 (2005).
- Qyang, Y. *et al.* The renewal and differentiation of Isl1<sup>+</sup> cardiovascular progenitors are controlled by a Wnt/ $\beta$ -catenin pathway. *Cell Stem Cell* **1**, 165–179 (2007).
- Moretti, A. *et al.* Multipotent embryonic Isl1<sup>+</sup> progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151–1165 (2006).
- Kawasaki, H. *et al.* Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. *Neuron* **28**, 31–40 (2000).
- Perrier, A. L. *et al.* Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **101**, 12543–12548 (2004).
- Wichterle, H., Lieberam, I., Porter, J. A. & Jessell, T. M. Directed differentiation of embryonic stem cells into motor neurons. *Cell* **110**, 385–397 (2002).
- Ying, Q. L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature Biotechnol.* **21**, 183–186 (2003).
- Li, X. J. *et al.* Specification of motoneurons from human embryonic stem cells. *Nature Biotechnol.* **23**, 215–221 (2005).
- Kattman, S. J., Huber, T. L. & Keller, G. M. Multipotent Flk-1<sup>+</sup> cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev. Cell* **11**, 723–732 (2006).
- Laflamme, M. A. *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnol.* **25**, 1015–1024 (2007).
- D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature Biotechnol.* **23**, 1534–1541 (2005).
- Warashina, M. *et al.* A synthetic small molecule that induces neuronal differentiation of adult hippocampal neural progenitor cells. *Angew. Chemie Int. Edn Engl.* **45**, 591–593 (2006).
- Diamandis, P. *et al.* Chemical genetics reveals a complex functional ground state of neural stem cells. *Nature Chem. Biol.* **3**, 268–273 (2007).
- Saxe, J. P. *et al.* A phenotypic small-molecule screen identifies an orphan ligand-receptor pair that regulates neural stem cell differentiation. *Chem. Biol.* **14**, 1019–1030 (2007).
- Hsieh, J., Nakashima, K., Kuwabara, T., Mejia, E. & Gage, F. H. Histone deacetylase inhibition-mediated neuronal differentiation of multipotent adult neural progenitor cells. *Proc. Natl Acad. Sci. USA* **101**, 16659–16664 (2004).
- Santarelli, L. *et al.* Requirement of hippocampal neurogenesis for the behavioral effects of antidepressants. *Science* **301**, 805–809 (2003).
- Hochedlinger, K. & Jaenisch, R. Nuclear reprogramming and pluripotency. *Nature* **441**, 1061–1067 (2006).
- Eggan, E. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428**, 44–49 (2004).
- Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416**, 545–548 (2002).
- Do, J. T. & Scholer, H. R. Nuclei of embryonic stem cells reprogram somatic cells. *Stem Cells* **22**, 941–949 (2004).
- Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).
- Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449**, 473–477 (2007).
- Cobaleda, C., Schebesta, A., Delogu, A. & Busslinger, M. Pax5: the guardian of B cell identity and function. *Nature Immunol.* **8**, 463–470 (2007).
- Kondo, T. & Raff, M. Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. *Science* **289**, 1754–1757 (2000).
- Shen, C. N., Slack, J. M. W. & Tosh, D. Molecular basis of transdifferentiation of pancreas to liver. *Nature Cell Biol.* **2**, 879–887 (2000).
- Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447**, 679–685 (2007).
- Byrne, J. A. *et al.* Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* **450**, 497–502 (2007).
- Yamanaka, S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cell* **1**, 39–49 (2007).



55. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
56. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25**, 1177–1181 (2007).
57. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55–70 (2007).
58. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
59. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
60. Zhang, F., Pomerantz, J. H., Sen, G., Palermo, A. T., & Blau, H. M. Active tissue-specific DNA demethylation conferred by somatic cell nuclei in stable heterokaryons. *Proc. Natl Acad. Sci. USA* **104**, 4395–4400 (2007).
61. Horb, M. E., Shen, C. N., Tosh, D. & Slack, J. M. W. Experimental conversion of liver to pancreas. *Curr. Biol.* **13**, 105–115 (2003).
62. Engler, A. J., Sen, S., Sweeney, H. L. & Discher, D. E. Matrix elasticity directs stem cell lineage specification. *Cell* **126**, 677–689 (2006).
63. Chen, S. B. *et al.* Reversine increases the plasticity of lineage-committed mammalian cells. *Proc. Natl Acad. Sci. USA* **104**, 10482–10487 (2007).
64. North, T. E. *et al.* Prostaglandin E<sub>2</sub> regulates vertebrate haematopoietic stem cell homeostasis. *Nature* **447**, 1007–1011 (2007).  
This paper describes a chemical screen in zebrafish that led to the identification of PGE<sub>2</sub> as a potent regulator of vertebrate HSC homeostasis.
65. Zhang, Q. S. *et al.* Small-molecule synergist of the Wnt- $\beta$ -catenin signaling pathway. *Proc. Natl Acad. Sci. USA* **104**, 7444–7448 (2007).
66. Adams, G. B. *et al.* Therapeutic targeting of a stem cell niche. *Nature Biotechnol.* **25**, 238–243 (2007).  
This paper provides a proof-of-principle demonstration that targeting stem-cell niches *in vivo* by conventional therapeutics can enhance stem-cell function and is an attractive strategy for regenerative medicine.
67. Koprivica, V. *et al.* EGFR activation mediates inhibition of axon regeneration by myelin and chondroitin sulfate proteoglycans. *Science* **310**, 106–110 (2005).
68. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* doi:10.1038/nature06968 (in the press).

**Acknowledgements** We thank members of the Ding laboratory for stimulating work and discussions. S.D. is supported by funding from the Scripps Research Institute, the National Institutes of Health (grant numbers MH074404, HD053759, HL084295 and HD058110), the Juvenile Diabetes Research Foundation and the California Institute of Regenerative Medicine.

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to S.D. ([sding@scripps.edu](mailto:sding@scripps.edu)).



# Imaging stem-cell-driven regeneration in mammals

Timm Schroeder<sup>1</sup>

**The ability to observe biological processes continuously, instead of at discrete time points, holds great promise for the study of tissue regeneration. Ideally, single cells would be followed continuously within large tissue volumes (such as organs) over long periods of time. Technical limitations, however, preclude such studies. But, recently, there have been improvements in imaging technologies and biologically compatible labelling agents. Together with new insights into the molecular characteristics of stem cells, which are ultimately responsible for the regenerative potential of all tissues, researchers are now much closer to applying single-cell imaging approaches to research into regeneration and its clinical applications.**

In mammals, the cells of some tissues, such as the skin, the blood and the gut epithelia, are constantly dying in normal homeostatic conditions and are successfully replaced throughout life — a process known as regeneration. Other tissues, such as those of the central nervous system, have a much lower cellular turnover. Whereas some mammalian tissues (such as the blood) can be restored to their original state after injury, most tissues can only be repaired but not completely restored. Tissue regeneration and repair are both highly complex and dynamic processes. They involve the interaction of many cells over days or weeks, and the cell-fate decisions of these cells need to be tightly coordinated: that is, whether each cell survives, self-renews (reproduces itself), proliferates, migrates or differentiates into a cell of a specific lineage at a particular point in time. The correct type and number of cells need to be produced and delivered to the right place at the right time for regeneration to be successful.

Ultimately, the potential for long-lasting regeneration lies with the stem cells of a tissue. It is their cell-fate decisions, together with those of their progeny, that result in successful sustained regeneration — or its failure. Deciphering how these decisions are controlled at the molecular level is crucial for understanding how regeneration occurs in homeostatic conditions and how it can be manipulated to regenerate tissues in the clinic.

The best and most direct approach to analyse how the constituents of complex three-dimensional structures interact is to visualize them. Imaging technologies therefore have long been a driving force in biomedical science, as shown by the number of Nobel prizes awarded for their development: for phase-contrast microscopy (in 1953), electron microscopy (1986), scanning tunnelling microscopy (1986), nuclear magnetic resonance spectroscopy (2002) and magnetic resonance imaging (MRI) (2003). However, as is apparent from optical illusions, visual data can be misleading: the brain can filter and interpret information in a way that results in inappropriate conclusions being drawn. This is even more likely to occur when observing continuous processes at insufficient temporal resolution (for example, when watching a time-lapse movie in which the intervals between the individual frames are too long). In this case, the brain tries to fill in the missing time periods with the most likely interpretation, on the basis of previous experience. In this way, researchers tend to see what they expect to find. It is not possible to fully understand complex continuous processes, such as the regeneration of organs from stem cells, if analyses rely on individual snapshots of populations of cells, as is illustrated

in Fig. 1. To exclude alternative explanations of experimental data, constant observation is indispensable<sup>1</sup>.

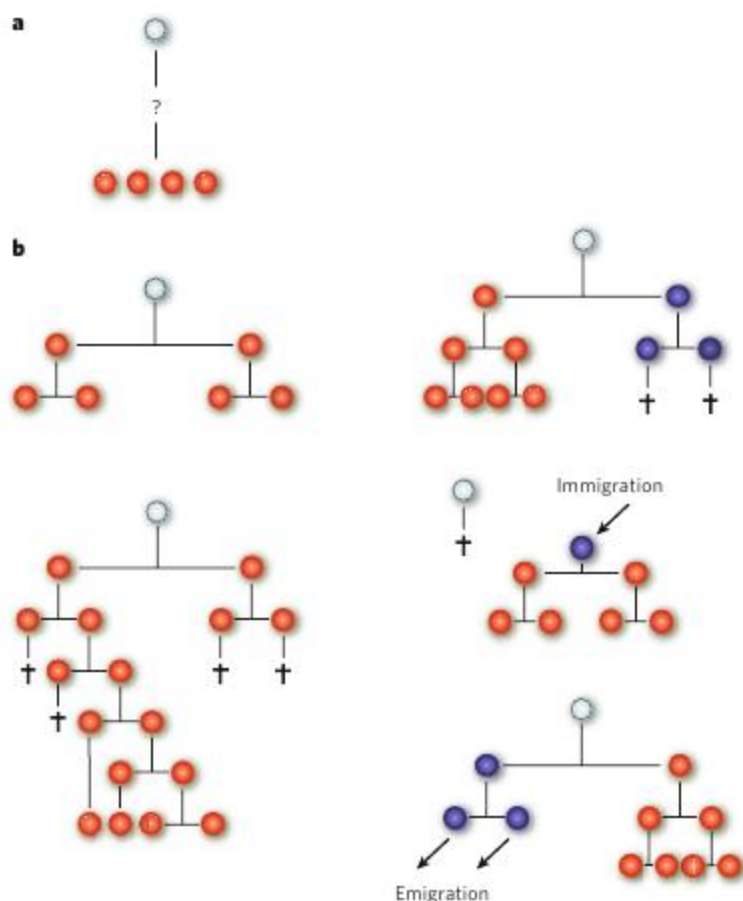
Observing the process of mammalian tissue regeneration in a continuous manner at the single-cell level would markedly improve knowledge about the underlying cellular mechanisms, and would lay the groundwork for a precise analysis of the molecular pathways that control this process and how they might be manipulated. For the purposes of this review, I define continuous imaging for following cells and their fates as time-lapse movies with individual frames collected over intervals of the order of a few minutes. This is in contrast to most studies on mammalian stem cells and their progeny, in which the cells are not analysed on the single-cell level and the time points are hours or days apart, even though the cell populations are not stationary (and can therefore move in the field of view). Decades-long debates in the field of stem-cell biology exemplify the confusion that can arise from misleading interpretations of discontinuous analyses. At present, there are controversies about many aspects, including the identity of the stem cells in most organs, the exact lineage potential of stem cells and progenitor cells, the proliferative activity of stem cells, the reasons for the incomplete reconstitution potential of transplanted stem cells, the possibility that stem cells have an asymmetrical mode of division (that is, that a single stem cell can both reproduce itself and differentiate into a progenitor cell during cell division), the transdifferentiation potential of adult stem cells, and the contribution of organized cellular movements to the generation of organs.

So far, there is no single imaging approach that is ideal for the continuous observation of all of the relevant aspects of stem-cell-driven regeneration. Instead, for each research question, optimal combinations of the individual cell system (that is, the tissue and its stem cells), the labelling approach for measuring the properties of interest, and the imaging modality for the detection of the labelled cells need to be found. Recent developments are, however, bringing researchers close to being able to observe mammalian stem-cell-driven regeneration continuously and comprehensively. The constant development of new imaging machinery, the adaptation of this machinery for the long-term observation of living animals and live cells and, in particular, the huge recent increase in the number of new labelling technologies and in the understanding of the molecular properties of stem cells are now allowing analyses that were previously impossible.

In this review, I focus on the current technical ability for continuously imaging mammalian stem-cell-driven regeneration, as well as its

<sup>1</sup>Institute of Stem Cell Research, Helmholtz Zentrum München – German Research Center for Environmental Health, Ingolstädter Landstrasse 1, D-85764 Neuherberg, Germany.





**Figure 1 | The need for continuous observation of dynamic cellular systems.**

Only continuous observation allows unequivocal interpretation of observed input and output data. **a**, A non-continuous observation. One cell is present at the first time point, and four are present at the next time point. But the intermediate steps are unclear. **b**, Some of the many possible explanations of the observation in **a** are shown. Each horizontal line represents a cell division, and crosses indicate cell death. Cells depicted in white are the initial cells. Red cells are of the cell type observed at the end of the observation period in **a** and **b**. Blue cells are of a different type; they might have been present in **a** but would have been missed in the absence of continuous observation.

limitations and future challenges. I discuss requirements for imaging, available imaging modalities and approaches to labelling, and I highlight several ways in which imaging approaches are being used to observe regeneration and repair in various mammalian tissues.

### Requirements for stem-cell imaging

Many of the properties of mammalian stem cells make them difficult to image.

Long-term *in vitro* culture could facilitate the continuous imaging of stem cells. But, for most stem cells, there are no physiological culture systems that allow long-term maintenance of self-renewal over sufficient periods. And, ultimately, it will be important to observe stem cells deep within live tissues.

For any given tissue, adult stem cells are rare and are surrounded by numerous other cells<sup>2–4</sup>. Therefore, unlike observing an aggregation of more homogeneous cells (such as some tumours), the detection of stem cells requires an imaging modality that has single-cell resolution. In addition, compared with imaging other cell types, the amount of usable data generated in each experiment is far less, owing to the low frequency of stem cells. Excellent examples of short-term single-cell observations *in vivo* are mouse models in which populations of identical T cells are transplanted and then observed after homing to the lymph nodes<sup>5</sup>. Because these T cells are identical and are numerous in the lymph nodes, the data that are collected for each of the many labelled T cells in a field of view can be used. By contrast, when imaging stem cells in a tissue, usually only one cell per observation is of interest. Therefore, generating enough data to carry out a reliable statistical analysis is extremely time-consuming, often to the extent

that it is not possible within a standard research funding period.

Another issue is that, with few exceptions<sup>2,4</sup>, researchers lack the knowledge for identifying live stem cells, or cells at other stages of maturation, within tissue. In cases in which this is possible, the differentiation state of the cell can be identified with reasonable reliability only by searching for a combination of molecular markers and by determining the position of the cell within the tissue<sup>2,4,6</sup>. Therefore, an imaging modality that allows multicolour labelling (a different colour for each marker) and detection must be used. In addition, labels that allow genetic marking (discussed later) are usually necessary.

The best — and, in most systems, the only reliable — definition of a stem cell is a cell that has the potential to differentiate into other cell types and can proliferate, survive and self-renew, all over long periods. To prove that a cell has these properties when using imaging experiments, live 'stem' cells (in a tissue or an animal) and at least some of their huge number of progeny need to be continuously observed at the single-cell level over long periods. Such continuous imaging would require techniques for tissue immobilization and long anaesthesia so that many cells can be followed for up to weeks at a time, but these techniques are not yet available. In general, continuous imaging results in unprecedented amounts of data, far exceeding the capabilities of typical imaging machinery and software. So the requirements for storing data and, more importantly, for comprehensively analysing these data cannot be met by the current tools used in stem-cell research.

Another consideration is that, in homeostatic conditions, adult stem cells usually divide much less frequently than most of their progeny. However, because cell division is the process by which all tissue cells are generated from stem cells, it is one of the most important cell-fate decisions to study. The slow rate of stem-cell division is therefore another factor that necessitates long-term imaging. Alternatively, many studies have been carried out after using injury to induce a high rate of stem-cell proliferation. Therefore, most knowledge about stem cells is based on observations of their behaviour in non-homeostatic conditions.

Furthermore, when imaging stem cells in homeostatic conditions, as well as after injury and particularly after transplantation, many stem cells and stem-cell progeny are highly migratory (see page 322) and can travel long distances. Therefore, continuous large-volume imaging of whole organs or organisms is required to observe them, and this imaging needs to be carried out at high temporal resolution (that is, with only very short intervals between each frame of the movie being constructed).

In conclusion, the imaging of stem-cell-driven regeneration requires an imaging technology that enables the behaviour, and ideally the molecular properties, of numerous cells to be observed continuously over many days, with high temporal resolution, at the single-cell level and in positions that are deep in large volumes of live tissue.

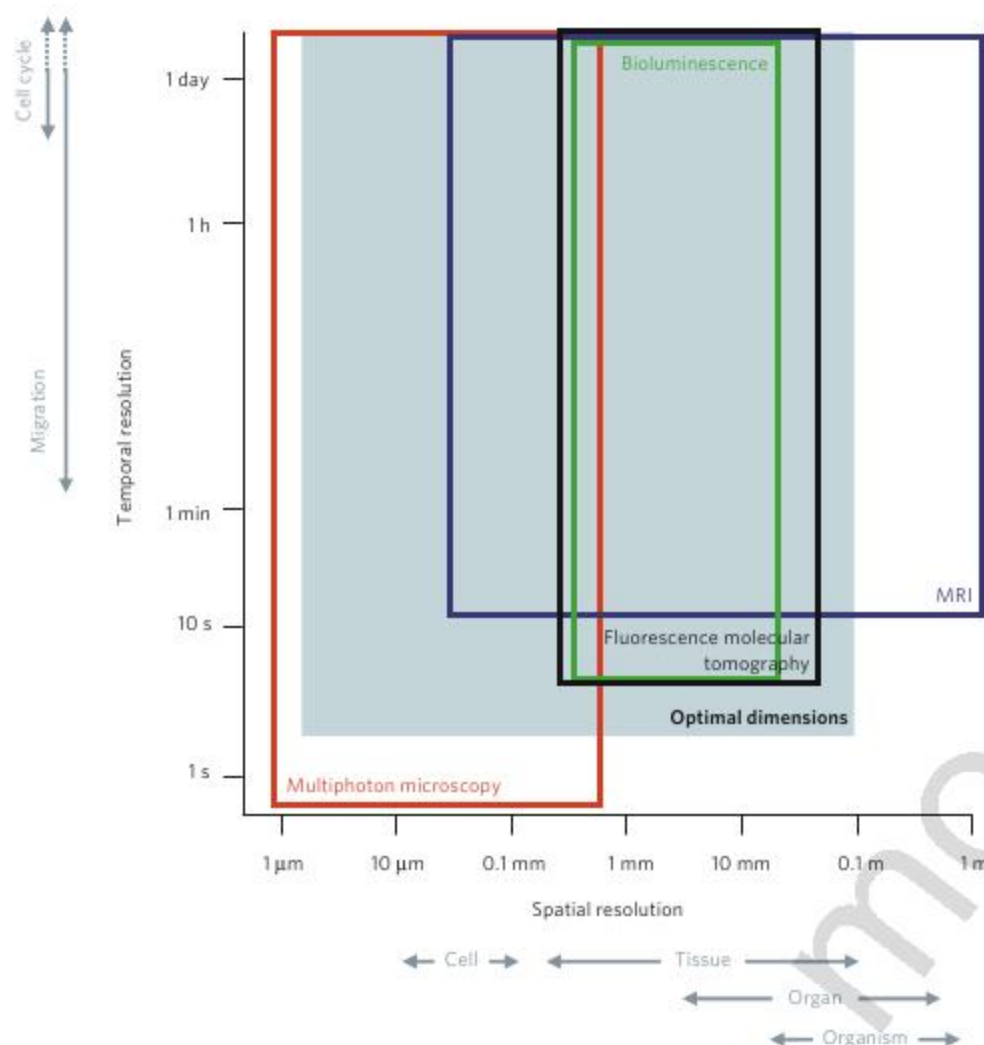
### Imaging modalities for regeneration research

All of the requirements mentioned in the previous section can be met individually by various imaging modalities<sup>7–9</sup>, but there is no single modality that allows a comprehensive approach to stem-cell imaging (Fig. 2). In this section, the most relevant imaging modalities for regeneration research are discussed: planar and tomographic fluorescence imaging (in this case wide-field fluorescence microscopy, confocal fluorescence microscopy, multiphoton fluorescence microscopy and fluorescence molecular tomography), bioluminescence imaging and MRI. The basic technical concepts underlying these modalities are described in Boxes 1 and 2.

The most important imaging modalities for live-cell imaging at single-cell resolution use light sources to excite fluorescent molecules (or fluorophores) that are attached to the cells of interest within the tissue (discussed later) and then detect the emitted fluorescent light<sup>9,10</sup>. These modalities include wide-field fluorescence microscopy, confocal fluorescence microscopy and multiphoton fluorescence microscopy. The necessary microscopes are comparatively easy to operate (easier than for fluorescence molecular tomography or MRI) and inexpensive, and are therefore widely available. The main limitations of these modalities are limited imaging depth in tissue and cell death as a result of phototoxicity, although new microscopes are continually being designed to reduce these problems<sup>11–13</sup>.

With fluorescence molecular tomography, live tissue can be visualized





**Figure 2 | The spatiotemporal resolution of different imaging modalities.** The filled grey rectangle indicates the optimal dimensions for the imaging of regeneration. The dimensions of imaging modalities applicable to regeneration research are overlaid. The lower horizontal lines of each box depict the highest temporal resolution, and the vertical lines on the left depict the highest spatial resolution. The upper horizontal lines depict the longest time periods that can be observed, and the vertical lines on the right depict the largest depths that can be observed. Typical values are shown; the precise performance of an imaging modality depends on the settings and the biological sample used. In terms of the cellular processes of interest, the cell cycle of a mammalian stem cell can last several months, and tissue regeneration from stem cells usually takes at least a few days. Observation periods longer than the indicated maxima are usually not limited by the method of signal detection but by factors relating to the incubation and immobilization of the biological sample, phototoxicity or data-handling capacities. MRI magnetic resonance imaging. (Figure adapted, with permission, from ref. 9.)

*in vivo* to a depth of several centimetres, allowing much larger volumes of tissue to be observed than by the other fluorescence imaging approaches that have been discussed<sup>8,14,15</sup>. Fluorescence molecular tomography allows extended continuous observation with high temporal resolution, but its spatial resolution is insufficient for observing single cells. However, the technology is developing fast, so improved imaging machinery and optimized modelling algorithms could soon make single-cell imaging possible. The development of new lasers, microscopes and fluorophores that allow long-wavelength (infrared) imaging will be important improvements for all four of these fluorescence imaging modalities (Box 2).

Bioluminescence imaging enables the observation of small animals over long periods without the need for a light source to excite fluorophores<sup>16</sup>. Cellular processes can be studied by following the expression of a genetically encoded luciferase enzyme, which cleaves substrates into forms that emit photons within live tissues. This approach can take advantage of the fact that luciferases from different organisms are specific for distinct substrates, which when cleaved emit photons of different wavelengths, so the photons can be distinguished, allowing the simultaneous use of more than one marker<sup>17</sup>. However, this system is far from single-cell resolution *in vivo*<sup>8</sup>: because the tissue penetration of the emitted light strongly depends on both the thickness of the tissue and the type of the tissue, and because no light source with a defined location is used, both quantification and exact reconstruction of the position of the bioluminescent light source are often not possible<sup>8</sup>.

MRI is among the least invasive of available imaging technologies<sup>7</sup>. Expensive new experimental machines have almost enough spatial and temporal resolution for continuous single-cell imaging. A major current limitation is the lack of genetically encodable labels with high signal strengths. In addition, the cost and complexity of the experimental machines with the strongest magnetic fields means that they are not widely available for researching cell-based therapies. The development of new labelling materials, together with technical improvements that further increase the spatial and temporal resolution and reduce the cost, could

make MRI the method of choice for tracking the behaviour of individual cells in large volumes<sup>18,19</sup>.

All individual imaging modalities have limitations in the spatial and/or temporal resolution or the time span and/or volume that can be observed in a single experiment (Fig. 2). In addition to improvements in the individual technologies, the desired specifications might also be met by using these methods simultaneously to combine their strengths<sup>8,18,20,21</sup>. For example, MRI can provide anatomical information about unstained tissues that would not be detectable by fluorescence imaging, while multiple molecular properties (labelled in different colours) can simultaneously be detected by fluorescence imaging of individual cells within this tissue<sup>22</sup>. Such multimodal imaging approaches are still in development but might soon overcome the current limitations in the imaging of regenerative processes.

### Labels for measuring cellular properties

Irrespective of the imaging modality used, the detection of individual cells and their functional properties is usually not possible without staining. Suitable agents for labelling the cellular property of interest are crucial, and a lack of these agents often limits experimental strategies. Typical requirements for these agents include long-term biological compatibility, brightness, stability and the possibility to distinguish several markers simultaneously. Numerous agents for molecular labelling are available<sup>23</sup>, but their usefulness depends on the cellular property of interest (Table 1).

To observe homeostatic regeneration experimentally, stem cells and their progeny at defined maturation stages must be labelled by using a genetic strategy, because direct manipulation would be invasive and is precluded by both the lack of accessibility of these cells and the difficulty in identifying them in live tissue.

The most important genetically encoded labels are fluorescent proteins. Since the first description of green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* more than 40 years ago<sup>24</sup>, many fluorescent proteins have been discovered or generated and have been progressively modified



**Box 1 | Selection of fluorophores based on relevant optical properties**

The following terms describe some basic parameters that determine the optical properties of fluorescent molecules (or fluorophores). In addition to biocompatibility and size (Table 1), these parameters are the most important when selecting fluorophores for biological imaging.

- **Fluorescence:** absorption of light of a particular wavelength and emission of light of a longer wavelength.
- **Stokes shift:** the difference between the excitation wavelength and emission wavelength of a fluorophore.
- **Extinction coefficient:** the efficiency of a fluorophore at absorbing photons at the excitation wavelength.
- **Quantum efficiency:** the ratio of the number of photons emitted by a fluorophore to the number of photons absorbed.
- **Photobleaching:** the conversion of a fluorophore into a non-fluorescent state by the absorbed photons.
- **Brightness:** the combination of the extinction coefficient, the quantum efficiency and photobleaching affects the brightness of a fluorophore.
- **Tissue penetrance:** the distance that light can travel through tissue without being scattered outside its original path.

To detect a fluorophore in a biological sample, the most important factor is the signal-to-noise ratio. The detectable signal depends on the brightness of the fluorophore and on the tissue penetrance of both the excitation and emission light. Also, in tissue, high background noise can result from the autofluorescence of biological molecules. Tissue penetrance improves with light of longer wavelengths, at which autofluorescence also contributes less to background noise. Therefore, for imaging deep within tissues, it is preferable to use fluorophores that absorb and emit light with a long wavelength (that is, light in the near-infrared region of the spectrum or longer-wavelength light).

to improve their imaging properties<sup>25,26</sup>. The simultaneous use of several distinguishable fluorescent proteins has become feasible through the generation of new fluorescent proteins with a variety of spectral properties<sup>26–29</sup>, and the generation of fluorescent proteins targeted to a variety of subcellular locations has further increased the combinatorial possibilities<sup>30</sup>. Nevertheless, there are still many improvements that could be made. For example, researchers are eagerly awaiting agents with bright near-infrared spectra<sup>29</sup>, which will result in greater tissue penetration and reduced tissue autofluorescence, thereby improving *in vivo* detection (Box 2).

When using new forms of fluorescent proteins, their biological compatibility must be carefully evaluated in the cells of interest. In many studies, an improved form of GFP known as enhanced GFP (eGFP) has successfully been used to label cells, including stem cells, without affecting the cells' behaviour<sup>4,31,32</sup>. In addition, a 'red mouse', in which all cells were labelled with red fluorescent protein with no obvious negative effect, was generated with the advent of a new label, DsRed-MST<sup>33</sup>. However, thorough analysis of the effects of the related label DsRed-Express later revealed that it had a negative influence on blood regeneration from haematopoietic stem cells (HSCs) compared with eGFP<sup>32</sup>. This study clearly shows the importance of carefully testing each new fluorescent protein for toxicity in the cell system to be analysed. Nevertheless, although the biological compatibility of fluorescent proteins depends on their expression level and on the cell type under study, the genes encoding eGFP<sup>31</sup>, enhanced cyan fluorescent protein and enhanced yellow fluorescent protein<sup>34–37</sup>, Venus<sup>38</sup>, DsRed-MST<sup>33</sup> and mRFP1 (monomeric red fluorescent protein 1)<sup>35,37,39</sup> have been expressed in stem cells or in whole mammals without an obvious change in phenotype, so these labels are highly likely to be non-toxic. It is important to note that animal models using unmodified eGFP as a marker are likely to be of limited use for future studies. The simultaneous marking of several genes of interest is crucial, for example, for the identification of stem cells or their progeny at specific maturation stages. This is not possible if all animal and cell lines are marked with the same fluorescent protein, yet most existing lines are marked by eGFP. When generating new animal lines, researchers should therefore make use of novel fluorescent proteins, which are likely to make these lines usable in combination with other labelled animal lines<sup>26,33,35–37</sup>.

In addition to fluorescent proteins, several other types of genetically encoded label are available for different imaging modalities, such as MRI. An elegant way to tag proteins for labelling is the genetic incorporation of small peptide sequences, which can then be detected by direct<sup>23,40</sup> or indirect<sup>41,42</sup> binding to fluorophores or to contrast agents for MRI. The small size of these peptide tags is less likely to result in steric hindrance in the targeted protein than when using a fluorescent protein, and the possibility of changing the fluorophore allows greater flexibility of complex experimental strategies. The production of particular enzymes can be visualized with high sensitivity *in vitro* by adding substrates that are converted into specific fluorescent or bioluminescent products<sup>21,43–45</sup>. Alternatively, for MRI studies, incorporation of the gene encoding ferritin into cells of interest can increase the contrast<sup>46</sup>. The number of genetically encoded contrast agents in development for multimodal imaging is also growing<sup>20,21,44</sup>.

The labelling of cells in this way, by genetic modification, is not safe for clinical therapy. Non-genetic labelling of cells could be achieved *in situ* by microinjection or, much more efficiently, *ex vivo* after purification (which then requires transplantation) (see page 322). The latter approach allows the delivery of almost any biologically compatible contrast agent (for example, organic fluorophores, quantum dots, and contrast labels for MRI), which is an advantage in that these agents have superior properties to genetic labels<sup>7,23,47,48</sup> (Table 1). However, the concentration of a non-genetic label is diluted with each cell division, so only the cells that have been labelled and their first progeny after transplantation can be visualized. Moreover, changes in cell states, such as differentiation, cannot usually be detected; such studies require genetic labelling. In addition, purifying and labelling cells *ex vivo* is invasive and can interfere with the normal processes of homeostatic regeneration that are being studied. Nevertheless, the tracking of manipulated and transplanted cells is of great interest for research into regeneration and repair, for example, to monitor homing to sites of injury.

Another type of non-genetic label is quantum dots, which are semiconductor nanocrystals with a biologically compatible coating. Quantum dots are extremely bright, are photostable and have size-dependent emission spectra that range from infrared to blue<sup>47</sup>, making them ideal for simultaneous multicolour imaging. However, the short wavelength of light required for the excitation of quantum dots means that their use in deep-tissue imaging is limited<sup>23</sup>. Quantum dots are still a relatively new technology but are increasingly being used by researchers. Recent reports strongly underline their potential for use in live-cell imaging<sup>49–51</sup>, and the first approaches to genetic labelling using quantum dots have recently been developed<sup>52</sup>.

**Analysis of imaging data**

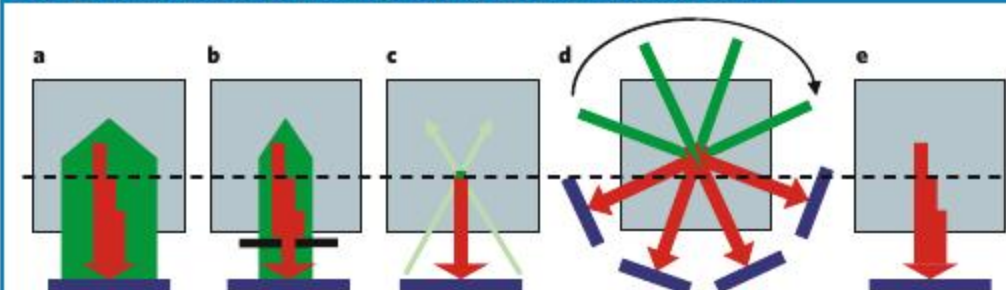
In addition to using the appropriate imaging modality and labelling strategy, it is important to consider how the data that are generated by an imaging study will be processed. Imaging experiments can be automated to generate large amounts of data<sup>53</sup>, but analysis of these data is the major bottleneck for high-throughput strategies. When searching for novel phenomena, the available tools for automated analysis are insufficient, and humans are currently better at recognizing patterns in two-dimensional and three-dimensional images. Therefore, with the need for human input at this level, data analysis is slow. So it will be crucial to develop better pattern-recognition algorithms for reliable automated detection and tracking of cells (and their behaviour) in live tissue<sup>54</sup>. In addition, the development of algorithms that can evaluate multidimensional data statistically and are adapted for the needs of stem-cell biology and regenerative medicine will be an important advance<sup>55,55</sup>. It is clear therefore that, although it is often regarded as a secondary problem, the availability of suitable data-analysis approaches is often a limiting factor in imaging studies.

**Imaging of different regenerating tissues**

The current possibilities and limitations of live-cell imaging in regeneration research can be illustrated by some of the studies that have been carried out in various organ systems: the cardiovascular system, skeletal muscle, the skin, the haematopoietic system and the nervous system.

Numerous imaging studies have analysed the behaviour of cells that have been transplanted to regenerate ischaemic heart muscle (see



**Box 2 | Technical aspects of imaging modalities for regeneration research**

The technical aspects of six imaging modalities that are commonly used in studies of regeneration — wide-field fluorescence microscopy, confocal fluorescence microscopy, multiphoton fluorescence microscopy, fluorescence molecular tomography, bioluminescence imaging and magnetic resonance imaging (MRI) — are discussed here.

The light paths in these imaging modalities (except for MRI) are shown in the figure. The imaged specimen is represented as a grey square, and excitation (green) and emission (red) light paths are depicted as arrows. The light-detector planes are shown as blue lines, and the focal plane is shown as dashed line.

In wide-field fluorescence microscopy (see figure, panel **a**), the entire field of vision is illuminated at the full depth of the sample, and all of the light that emits from the sample — from within and outside the focal plane — is projected through an objective onto one detection plane. This is faster, more sensitive and less phototoxic (that is, causes less cell death) than confocal fluorescence microscopy approaches, and it allows live cells to be imaged with subcellular resolution, even over weeks<sup>9,10,67</sup>. However, cells can be followed to a depth of only a few cell diameters. This method is therefore ideally suited to long-term two-dimensional cell imaging of *ex vivo* samples.

In confocal fluorescence microscopy (see figure, panel **b**), monochromatic laser light is usually used for excitation, and the light emitted from only a single focal point is detected. Two-dimensional images are compiled by sequentially scanning all individual points in one plane. This allows optical sectioning and better volume imaging than wide-field fluorescence microscopy<sup>75</sup>. But, for each of the scanned points in one plane, the excitation laser beam also illuminates all points under and over the scanned focal point, so phototoxicity is high.

In multiphoton fluorescence microscopy (see figure, panel **c**), more than one laser beam is used to excite the fluorophore, but each light beam is of lower energy than that used in confocal fluorescence microscopy<sup>9,75</sup>. Only at the focal point does the laser light combine to yield the energy needed for fluorophore excitation. All other points in the tissue therefore receive less energy, reducing phototoxicity. In addition, because light of

a longer wavelength (which has less energy) is used, this imaging approach has a greater tissue penetration than confocal fluorescence microscopy, allowing imaging to a depth of many hundred micrometres. However, the volumes that can be visualized with this modality are still much too small to allow the observation of whole mammalian organs.

In fluorescence molecular tomography (see figure, panel **d**), light is propagated from a source through the tissue to a detector and is

sequentially analysed at several planar angles. By applying mathematical models for light propagation in tissue, data generated by this tomographic approach can be used to reconstruct reliably the position of fluorophores to a depth of several centimetres in a quantitative way<sup>8</sup>. The spatial resolution of this method is still too low at present for reliable detection of single cells (Fig. 2). But the theoretical resolution limits of this developing technology have not yet been reached, and improvements in imaging machinery and in modelling algorithms are soon likely to improve resolution almost to the cellular level.

In bioluminescence imaging (see figure, panel **e** for the planar approach), photons that are generated in tissue in the absence of excitation light — that is, photons emitted following the conversion of a substrate by, for example, genetically encoded luciferase — can be detected by planar or tomographic approaches<sup>16</sup>. Emitted light of long wavelengths (>600 nm) allows tissue penetration to a depth of more than 1 cm. However, it is not possible to localize the bioluminescent source at cellular resolution in tissue, and hundreds to thousands of cells are usually needed for detectable signals.

In MRI, atoms are aligned in a strong magnetic field and excited by radio waves, and then the electromagnetic signals emitted by these atoms are detected. Different types of atom have different absorption and emission spectra, and these spectra can be compiled to generate a contrast image<sup>7</sup>. In addition, paramagnetic contrast agents, which yield very strong signals, can be used to enhance the image contrast. MRI can be applied to large volumes in whole living organisms and has a central role in routine clinical imaging of large-volume soft tissues. The current drawback of this technology for experimental research is its limited spatial and temporal resolution. The resolution depends on the strength of the magnetic field and the availability of labels that yield high signal strengths. MRI machines used for clinical imaging typically operate with field strengths of up to 3 tesla, but expensive experimental machines can operate at more than 20 tesla, allowing imaging of whole live organs with almost enough spatial and temporal resolution to follow regenerative processes continuously at the single-cell level.

page 322). In these studies, mesenchymal stem cells, bone-marrow cells and embryonic stem (ES)-cell-derived cells were labelled with imaging agents, transplanted into animals and followed by various imaging modalities<sup>22,56–58</sup>. The migration of these cells to the heart and the survival of their progeny was observed. However, the long-term behaviour of individual cells and whether they can regenerate damaged tissues *in vivo* has not yet been studied in detail.

The cellular mechanisms of skeletal muscle regeneration have recently gained attention<sup>59</sup>, and studies using continuous fluorescence imaging have identified skeletal muscle stem cells and pinpointed their mode of division<sup>60</sup>. In addition, continuous fluorescence imaging *ex vivo* has recently confirmed that the original DNA and the newly synthesized DNA are asymmetrically inherited during the division of skeletal muscle stem cells *in vitro*<sup>61</sup>.

The skin is almost an ideal tissue for imaging. It is easier to access *in vivo* than other tissues, and the high turnover rates of skin cells make it possible to observe the complete development of stem cells into mature progeny in a short period. Skin cells are also relatively stationary and are organized in a simple three-dimensional arrangement compared with other tissues. In addition, genetic markers of the various maturation states of skin cells are better defined than for most other stem-cell systems<sup>62</sup>, and

there are mouse lines that allow the fluorescent labelling of epithelial stem cells in the hair follicle<sup>4</sup> or the clonal progeny of individual stem cells in the epidermis<sup>63</sup>. It is therefore surprising that the skin has not been used extensively to observe the homeostatic behaviour of stem cells, and it is likely to become the subject of *in vivo* stem-cell imaging studies soon.

Haematopoiesis is arguably among the most difficult of the mammalian stem-cell systems to image continuously. Blood cells are highly migratory and reside in liquid tissues without known three-dimensional tissue organization, and most lineages of blood cell cannot reliably be distinguished by morphology either *in vivo* or *in vitro*. In homeostatic conditions, HSCs divide infrequently, and non-invasive long-term imaging is more challenging in the bone marrow, where HSCs reside, than in many other tissues. Even so, the continuous observation of homeostatic or transplanted haematopoietic cells has been attempted more often than for cells of many other tissues<sup>5,16,64–66</sup>. Haematopoiesis is better understood than other stem-cell systems and has huge clinical importance, but despite intensive research in the past decade, many basic questions are still unresolved<sup>67</sup>. For this reason, the importance of continuous observation of cellular behaviour soon became clear and led to imaging attempts earlier than in other systems. For example, bioluminescence imaging was used to visualize the progeny of transplanted HSCs in various anatomical locations



**Table 1 | Properties of typical fluorophores**

Fluorophore type	Brightness	Direct genetic encoding	Long-term biological compatibility	Tissue penetrance	Non-invasive <i>in vivo</i> labelling	Multi-colour labelling
Fluorescent proteins	+	++	+++	+/-	++	+
Peptide tags (indirect labelling)	++	++	ND	++	+/-	+
Organic fluorophores	++	-	++	++	-	++
Quantum dots	+++	-	ND	++*	-	+++

ND, not determined; -, very difficult; +/-, possible; +, good; ++, very good; +++, excellent. \*Can be excellent for emitted light but is limited for excitation light with short wavelengths.

in recipient animals (such as the bone marrow of the legs, the vertebrae and the ribs)<sup>16</sup>, but the resolution was many orders of magnitude lower than single-cell resolution. Continuous observation of haematopoietic progenitor cells in the bone marrow was achieved at single-cell resolution by using multiphoton fluorescence microscopy<sup>64,66</sup>. This modality was also used to study the homing of normal and malignant haematopoietic progenitor cells in the bone marrow and to characterize the specialized niches of these cells<sup>65</sup>.

Considerable efforts have been made to reduce the number of molecular markers required for reliable prospective identification of HSCs in homeostatic conditions<sup>3</sup> or haematopoietic cells of specific maturation states<sup>68</sup>. So far, the most successful approach has enabled homeostatic HSCs to be identified with up to 50% reliability using a combination of markers but only two colours<sup>6</sup>. Further identification of individual markers that allow the unequivocal prospective identification of different maturation stages of haematopoietic cells will be crucial for generating transgenic animals that can be used to observe haematopoiesis in the bone marrow continuously.

In contrast to most tissues in adults, the central nervous system has a low regenerative activity, and neural stem cells reside in regions of the adult brain that are difficult to access by most imaging modalities<sup>69</sup>. Therefore, an important model system for studying neural regeneration is the developing central nervous system in the embryo. Slice preparations of developing central-nervous-system tissue can be imaged continuously for a limited time *in vitro*, and such studies have provided direct proof of the neurogenic potential of glial cells and have identified the division modes that generate neurons<sup>70</sup>. In addition, neural-cell migration has been analysed in live embryonic brains *in utero* by using multiphoton microscopy<sup>71</sup>, opening up the possibility of studying neurogenesis in the intact developing central nervous system *in vivo*. Recent studies using various time-lapse imaging modalities have analysed both homeostatic regeneration and injury-induced repair processes in the adult rodent brain and the human brain<sup>72-74</sup>. In the future, an important goal will be to follow individual cells over a period of days during regeneration, with read-outs for the changing differentiation states.

Finally, ES cells hold great promise for regenerative medicine. ES cells can be maintained indefinitely in an undifferentiated state *in vitro*, are highly proliferative and have the potential to differentiate into every cell type in the body (see page 306). In addition, molecular markers of many of their differentiation states have been identified. For these reasons, ES cells are ideal for continuous imaging *in vitro*. Surprisingly, they have not yet been used extensively for the long-term study of differentiation events. But human ES cells, in particular, are likely to increase in importance, forming the basis of a clinically relevant, physiological *in vitro* model of regeneration that can be used for continuous observation at the single-cell level.

## Future developments

The ideal situation would be to observe unstained whole organisms over a period of weeks, with a spatial resolution at the molecular level and a temporal resolution of milliseconds. The technology should be applicable to humans receiving therapy and suitable for automated high-throughput data collection, with subsequent unsupervised comprehensive analysis of these data. Clearly, this goal is far from reach, but the development

of even a single new technology could lead to a breakthrough in how regeneration is understood.

In clinical studies, the imaging of degenerating or regenerating heart, skin and cartilage would provide important insights into how cell populations migrate and turn over, how stem cells and progenitor cells contribute to tissue development, and how regenerative processes and inflammatory processes interact. Most fluorescence imaging approaches, however, are not safe enough for studying humans and have observation volumes that are too limited to be useful in humans. Therefore, MRI is likely to remain the most important imaging modality for observing regenerative processes in humans. Because improving the sensitivity and image resolution of MRI is expensive (and only small linear improvements in resolution are possible), the development of new biologically compatible agents that provide strong contrast holds the greatest promise for improving the images obtained by MRI of the human body, allowing the detection of smaller populations of cells.

Using animal models and extrapolating the findings to humans might therefore be the best way to observe specific sites, cell populations or time windows that are of interest during regenerative processes in humans, as well as to reduce the requirements for the imaging modalities that are used. The development of small optical imaging sensors (such as implantable miniaturized microscopy systems) that allow endoscopic access or even transient implantation at sites of interest could help to facilitate the use of shorter-range but higher-resolution optical modalities. These sensors could also help to reduce the problems associated with long anaesthesia and possibly with tissue immobilization, both of which are required for long-term observations with current imaging modalities. In terms of the modality, mesoscopic tomography<sup>8</sup> (which allows observations to be made in the millimetre to centimetre range) holds the greatest promise for imaging in animal models: there is still room for substantial technical improvement not only to the machinery but also to the modelling algorithms underlying this relatively new technology. The spatial and temporal resolution and the volumes that are observable with this modality are likely to increase soon, together with improvements in the ability to observe light of several wavelengths simultaneously, possibly enabling whole live organs to be observed continuously at single-cell resolution.

The greatest potential for optimizing imaging approaches for regeneration research probably lies in applying new insights from stem-cell biology and well-defined animal models of regeneration. The identification of additional molecules that mark specific cellular properties or maturation states during stem-cell differentiation would enable these properties to be labelled with distinct agents of various colours (or spectral properties), which are constantly being improved. Invertebrate and zebrafish model systems have superior properties for imaging as a result of their small size, low opacity and fast development from embryo to adult. Despite anatomical differences, insights into dynamic cellular behaviour that are gained using these models could help to focus imaging efforts in, for example, mouse models, with the ultimate goal being to confirm the findings about regenerative mechanisms by subsequently studying humans.

The imaging of regenerative processes involves researchers from many disciplines and is at constant risk of being delayed by the differences in interests, technical language, funding and philosophy in the relevant areas of academic and industry research. Despite this, the potential benefits of regenerative medicine should be a strong motivation to continuously improve imaging strategies that will enable stem-cell-driven regeneration in mammals to be completely understood — simply by letting researchers watch it happen.

1. Yamanaka, Y., Tamplin, O. J., Beckers, A., Gossler, A. & Rossant, J. Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. *Dev. Cell* **13**, 884–896 (2007).
2. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
3. Osawa, M., Hanada, K., Hamada, H. & Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. *Science* **273**, 242–245 (1996).
4. Tumber, T. et al. Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004). This paper was the first description of non-invasive labelling of mammalian homeostatic stem cells with a fluorescent protein.
5. Sumen, C., Mempel, T. R., Mazo, I. B. & von Andrian, U. H. Intravital microscopy: visualizing



- immunity in context. *Immunity* **21**, 315–329 (2004).
6. Kiel, M. J. et al. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
  7. Jacobs, R. E., Papan, C., Ruffins, S., Tyska, J. M. & Fraser, S. E. MRI: volumetric imaging for vital imaging and atlas construction. *Nature Rev. Mol. Cell Biol.* **4**, SS10–SS16 (2003).
  8. Ntziachristos, V., Ripoll, J., Wang, L. V. & Weissleder, R. Looking and listening to light: the evolution of whole-body photonic imaging. *Nature Biotechnol.* **23**, 313–320 (2005). **This excellent review compares optical tomography with other optical imaging modalities.**
  9. Tsien, R. Y. Imagining imaging's future. *Nature Rev. Mol. Cell Biol.* **4**, SS16–SS21 (2003). **This review provides a concise overview of various imaging modalities.**
  10. Lichtman, J. W. & Conchello, J. A. Fluorescence microscopy. *Nature Methods* **2**, 910–919 (2005).
  11. Campagnola, P. J. & Loew, L. M. Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nature Biotechnol.* **21**, 1356–1360 (2003).
  12. Hoebe, R. A. et al. Controlled light-exposure microscopy reduces photobleaching and phototoxicity in fluorescence live-cell imaging. *Nature Biotechnol.* **25**, 249–253 (2007).
  13. Verveer, P. J. et al. High-resolution three-dimensional imaging of large specimens with light sheet-based microscopy. *Nature Methods* **4**, 311–313 (2007).
  14. Vinegoni, C., Pitsouli, C., Razansky, D., Perrimon, N. & Ntziachristos, V. In vivo imaging of *Drosophila melanogaster* pupae with mesoscopic fluorescence tomography. *Nature Methods* **5**, 45–47 (2008).
  15. Zacharakis, G. et al. Volumetric tomography of fluorescent proteins through small animals in vivo. *Proc. Natl Acad. Sci. USA* **102**, 18252–18257 (2005).
  16. Cao, Y. A. et al. Shifting foci of hematopoiesis during reconstitution from single stem cells. *Proc. Natl Acad. Sci. USA* **101**, 221–226 (2004).
  17. Wood, K. V., Lam, Y. A., Seliger, H. H. & McElroy, W. D. Complementary DNA coding click beetle luciferases can elicit bioluminescence of different colors. *Science* **244**, 700–702 (1989).
  18. Sumner, J. P., Conroy, R., Shapiro, E. M., Moreland, J. & Koretsky, A. P. Delivery of fluorescent probes using iron oxide particles as carriers enables in-vivo labeling of migrating neural precursors for magnetic resonance imaging and optical imaging. *J. Biomed. Opt.* **12**, 051504 (2007).
  19. Shapiro, E. M., Sharer, K., Skrtic, S. & Koretsky, A. P. In vivo detection of single cells by MRI. *Magn. Reson. Med.* **55**, 242–249 (2006).
  20. Panomarev, V. et al. A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur. J. Nucl. Med. Mol. Imaging* **31**, 740–751 (2004).
  21. Ray, P., Tsien, R. & Gambhir, S. S. Construction and validation of improved triple fusion reporter gene vectors for molecular imaging of living subjects. *Cancer Res.* **67**, 3085–3093 (2007).
  22. Sosnovik, D. E. et al. Fluorescence tomography and magnetic resonance imaging of myocardial macrophage infiltration in infarcted myocardium in vivo. *Circulation* **115**, 1384–1391 (2007).
  23. Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217–224 (2006). **This review provides a comprehensive description of the various possibilities for fluorescent labelling of cells.**
  24. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusa, *Aequorea*. *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).
  25. Matz, M. V., Lukyanov, K. A. & Lukyanov, S. A. Family of the green fluorescent protein: journey to the end of the rainbow. *Bioessays* **24**, 953–959 (2002).
  26. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nature Methods* **2**, 905–909 (2005).
  27. Ai, H. W., Shaner, N. C., Cheng, Z., Tsien, R. Y. & Campbell, R. E. Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **46**, 5904–5910 (2007).
  28. Nagai, T. et al. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnol.* **20**, 87–90 (2002).
  29. Shcherbo, D. et al. Bright far-red fluorescent protein for whole-body imaging. *Nature Methods* **4**, 741–746 (2007).
  30. Okita, C., Sato, M. & Schroeder, T. Generation of optimized yellow and red fluorescent proteins with distinct subcellular localization. *Biotechniques* **36**, 418–422 (2004).
  31. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
  32. Tao, W. et al. Enhanced green fluorescent protein is a nearly ideal long-term expression tracer for hematopoietic stem cells, whereas DsRed-Express fluorescent protein is not. *Stem Cells* **25**, 670–678 (2007). **This paper describes a comprehensive quantitative study of the effects of DsRed-Express on the long-term behaviour of stem cells.**
  33. Vintersten, K. et al. Mouse in red: red fluorescent protein expression in mouse ES cells, embryos, and adult animals. *Genesis* **40**, 241–246 (2004).
  34. Hadjantonakis, A. K., Macmaster, S. & Nagy, A. Embryonic stem cells and mice expressing different GFP variants for multiple non-invasive reporter usage within a single animal. *BMC Biotechnol.* **2**, 11 (2002).
  35. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
  36. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
  37. Ueno, H. & Weissman, I. L. Clonal analysis of mouse development reveals a polyclonal origin for yolk sac blood islands. *Dev. Cell* **11**, 519–533 (2006).
  38. Takayanagi, S. et al. Genetic marking of hematopoietic stem and endothelial cells: identification of the *Tmtsp* gene encoding a novel cell surface protein with the thrombospondin-1 domain. *Blood* **107**, 4317–4325 (2006).
  39. Long, J. Z., Lackan, C. S. & Hadjantonakis, A. K. Genetic and spectrally distinct in vivo imaging: embryonic stem cells and mice with widespread expression of a monomeric red fluorescent protein. *BMC Biotechnol.* **5**, 20 (2005).
  40. Kelly, K. A., Carson, J., McCarthy, J. R. & Weissleder, R. Novel peptide sequence ('IQ-tag') with high affinity for NIR fluorochromes allows protein and cell specific labeling for in vivo imaging. *PLoS ONE* **2**, e665 (2007).
  41. Fernandez-Suarez, M. et al. Redirecting lipase for cell surface protein labeling with small-molecule probes. *Nature Biotechnol.* **25**, 1483–1487 (2007).
  42. Tannous, B. A. et al. Metabolic biotinylation of cell surface receptors for in vivo imaging. *Nature Methods* **3**, 391–396 (2006).
  43. Gao, W., Xing, B., Tsien, R. Y. & Rao, J. Novel fluorogenic substrates for imaging  $\beta$ -lactamase gene expression. *J. Am. Chem. Soc.* **125**, 11146–11147 (2003).
  44. Josseland, V., Texier-Nogues, I., Huber, P., Favrot, M. C. & Coll, J. L. Non-invasive in vivo optical imaging of the *locZ* and *luc* gene expression in mice. *Gene Ther.* **14**, 1587–1593 (2007).
  45. Naik, S. & Pivnicka-Worms, D. Real-time imaging of  $\beta$ -catenin dynamics in cells and living mice. *Proc. Natl Acad. Sci. USA* **104**, 17465–17470 (2007).
  46. Cohen, B. et al. MRI detection of transcriptional regulation of gene expression in transgenic mice. *Nature Med.* **13**, 498–503 (2007).
  47. Alivisatos, P. The use of nanocrystals in biological detection. *Nature Biotechnol.* **22**, 47–52 (2004).
  48. Swirski, F. K. et al. A near-infrared cell tracker reagent for multiscope in vivo imaging and quantification of leukocyte immune responses. *PLoS ONE* **2**, e1075 (2007).
  49. Lidke, D. S. et al. Quantum dot ligands provide new insights into erbB/HER receptor-mediated signal transduction. *Nature Biotechnol.* **22**, 198–203 (2004).
  50. Slotkin, J. R. et al. In vivo quantum dot labeling of mammalian stem and progenitor cells. *Dev. Dyn.* **236**, 3393–3401 (2007).
  51. Zhang, Q., Cao, Y. Q. & Tsien, R. Y. Quantum dots provide an optical signal specific to full collapse fusion of synaptic vesicles. *Proc. Natl Acad. Sci. USA* **104**, 17843–17848 (2007).
  52. Bonasio, R. et al. Specific and covalent labeling of a membrane protein with organic fluorochromes and quantum dots. *Proc. Natl Acad. Sci. USA* **104**, 14753–14758 (2007).
  53. Megason, S. G. & Fraser, S. E. Imaging in systems biology. *Cell* **130**, 784–795 (2007).
  54. Megason, S. G. & Fraser, S. E. Digitizing life at the level of the cell: high-performance laser-scanning microscopy and image analysis for in toto imaging of development. *Mech. Dev.* **120**, 1407–1420 (2003).
  55. Glauche, I., Cross, M., Loeffler, M. & Roeder, I. Lineage specification of hematopoietic stem cells: mathematical modeling and biological implications. *Stem Cells* **25**, 1791–1799 (2007).
  56. Ebert, S. N. et al. Noninvasive tracking of cardiac embryonic stem cells in vivo using magnetic resonance imaging techniques. *Stem Cells* **25**, 2936–2944 (2007).
  57. Laflamme, M. A. et al. Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnol.* **25**, 1015–1024 (2007).
  58. van Laake, L. W. et al. Monitoring of cell therapy and assessment of cardiac function using magnetic resonance imaging in a mouse model of myocardial infarction. *Nature Protocols* **2**, 2551–2567 (2007).
  59. Peault, B. et al. Stem and progenitor cells in skeletal muscle development, maintenance, and therapy. *Mol. Ther.* **15**, 867–877 (2007).
  60. Kuang, S., Kuroda, K., Le Grand, F. & Rudnicki, M. A. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell* **129**, 999–1010 (2007).
  61. Shinin, V., Gayraud-Morel, B., Gomes, D. & Tajbakhsh, S. Asymmetric division and cosegregation of template DNA strands in adult muscle satellite cells. *Nature Cell Biol.* **8**, 677–687 (2006).
  62. Fuchs, E. Scratching the surface of skin development. *Nature* **445**, 834–842 (2007).
  63. Ro, S. & Rannala, B. A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* **5**, 914–920 (2004).
  64. Junt, T. et al. Dynamic visualization of thrombopoiesis within bone marrow. *Science* **317**, 1767–1770 (2007).
  65. Sipkins, D. A. et al. In vivo imaging of specialized bone marrow endothelial microdomains for tumour engraftment. *Nature* **435**, 969–973 (2005). **This study is an excellent example of imaging of the bone marrow of a live mouse by using time-lapse multiphoton microscopy.**
  66. Suzuki, N. et al. Combinatorial Gata2 and Sca1 expression defines hematopoietic stem cells in the bone marrow niche. *Proc. Natl Acad. Sci. USA* **103**, 2202–2207 (2006).
  67. Schroeder, T. Tracking hematopoiesis at the single cell level. *Ann. NY Acad. Sci.* **1044**, 201–209 (2005).
  68. Heck, S. et al. Distinguishable live erythroid and myeloid cells in  $\beta$ -globin ECFP x lysozyme EGFP mice. *Blood* **101**, 903–906 (2003).
  69. Merkle, F. T., Mirzadeh, Z. & Alvarez-Buylla, A. Mosaic organization of neural stem cells in the adult brain. *Science* **317**, 381–384 (2007).
  70. Miyata, T., Kawaguchi, A., Okano, H. & Ogawa, M. Asymmetric inheritance of radial glial fibers by cortical neurons. *Neuron* **31**, 727–741 (2001).
  71. Yokota, Y. et al. Radial glial dependent and independent dynamics of interneuronal migration in the developing cerebral cortex. *PLoS ONE* **2**, e794 (2007).
  72. Gobel, W., Kampa, B. M. & Helmchen, F. Imaging cellular network dynamics in three dimensions using fast 3D laser scanning. *Nature Methods* **4**, 73–79 (2007).
  73. Manganas, L. N. et al. Magnetic resonance spectroscopy identifies neural progenitor cells in the live human brain. *Science* **318**, 980–985 (2007).
  74. Mizrahi, A. Dendritic development and plasticity of adult-born neurons in the mouse olfactory bulb. *Nature Neurosci.* **10**, 444–452 (2007).
  75. Conchello, J. A. & Lichtman, J. W. Optical sectioning microscopy. *Nature Methods* **2**, 920–931 (2005).

**Acknowledgements** I apologize to colleagues whose work could not be cited because of space constraints. I thank V. Ntziachristos, L. Godinho and M. Sixt for discussions, and M. A. Rieger, A. Ippen and E. Drew for critical reading of the manuscript. This work was financed in part by the German Research Foundation (DFG).

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). The author declares no competing financial interests. Correspondence should be addressed to the author ([tim.schroeder@helmholtz-muenchen.de](mailto:tim.schroeder@helmholtz-muenchen.de)).



## LETTERS

# True polar wander on Europa from global-scale small-circle depressions

Paul Schenk<sup>1</sup>, Isamu Matsuyama<sup>2</sup> & Francis Nimmo<sup>3</sup>

The tectonic patterns and stress history of Europa are exceedingly complex<sup>1</sup> and many large-scale features remain unexplained. True polar wander, involving reorientation of Europa's floating outer ice shell about the tidal axis with Jupiter, has been proposed<sup>2</sup> as a possible explanation for some of the features. This mechanism is possible if the icy shell is latitudinally variable in thickness and decoupled from the rocky interior. It would impose high stress levels on the shell, leading to predictable fracture patterns<sup>3</sup>. No satisfactory match to global-scale features has hitherto been found for polar wander stress patterns<sup>4</sup>. Here we describe broad arcuate troughs and depressions on Europa that do not fit other proposed stress mechanisms in their current position. Using imaging from three spacecraft, we have mapped two global-scale organized concentric antipodal sets of arcuate troughs up to hundreds of kilometres long and 300 m to ~1.5 km deep. An excellent match to these features is found with stresses caused by an episode of ~80° true polar wander. These depressions also appear to be geographically related to other large-scale bright and dark lineaments<sup>4,5</sup>, suggesting that many of Europa's tectonic patterns may also be related to true polar wander.

Three arcuate troughs 25–40 km across and ≥500 km long have been identified on Europa from imaging data obtained with Voyager, Galileo and New Horizons spacecraft (Figs 1, 2). Although no bounding walls, scarps or geologic discontinuities can be resolved in Galileo images, the observed topographic subsidence suggests that extension has occurred. Two troughs are observed on the leading hemisphere, forming opposing quadrants of a large, essentially perfectly circular, feature 2,350 km in diameter (Fig. 2). Together, these arcs, referred to as circle L, have a least-squares best-fit centre<sup>6</sup> of 10°S, 120° ± 1°W and spherical radius of 43.3°. Circle L is of a type known as a 'small circle', as its plane does not go through the centre of Europa. Neither trough deviates more than 0.3° from circularity (Supplementary Fig. 1). The circle is symmetric but incomplete; troughs are absent in the northwest and southeast sectors, despite suitable imaging (Figs 1, 2), and apparently did not form.

Only one arcuate trough segment was observed on the trailing hemisphere (Figs 1c, 2). This trough has a best-fit centre of 10°N, 300°W, antipodal to circle L and forming part of a second circular trough feature, circle T. By symmetry with circle L, we expect another segment near 270°W, but neither Voyager nor Galileo acquired low-Sun imaging in this region (Fig. 2). No trough has as yet been identified in New Horizons<sup>7</sup> imaging of this region, owing to the morphologic complexity of the region and low image resolution (~17 km per pixel).

Several shorter arcuate closed depressions 100–200 km long also occur parallel to, and in close association with, the long arcuate troughs (Figs 1, 2). These occur 48° to 52° from the respective best-fit centres in widely dispersed positions 100 to 300 km exterior

to the larger arcuate troughs. Two morphologies occur (Fig. 1): sets of two to three parallel short linear troughs <100 km apart, and isolated elongate bulbous-shaped steep-walled depressions (Fig. 3).

Limited stereo and photoclinometric topography (Supplementary Information) indicate relief across arcuate troughs of 300 ± 50 m, but relief across secondary basins is found to be 900–1,500 m. The latter are the deepest features known on Europa, with relief comparable to elevated plateaus elsewhere<sup>8–10</sup>, indicating a total dynamic range of relief on Europa of nearly 2 km. Despite their depths and pervasive deformation across most of Europa<sup>11</sup>, there are no indications of disruption, deformation or lateral displacement of pre-existing features across the floors of these depressions, nor are there indications of emplacement of new material on these depression floors (Fig. 3). Ridges and chaos on Europa either pre-date these depressions, or formed unaffected by their presence.

The observed troughs and secondary basins (Fig. 2) form two well-defined and antipodal sets of concentric small-circle depressions (SCDs)<sup>12</sup>, offset 10° from the equator and 30° from the present tidal axis (oriented towards Jupiter). SCD geography appears correlated with several major geologic features on Europa. A great circle oriented ~50° to the equator through the best-fit centres (Fig. 4) approximately bisects all three observed trough segments, and passes near and parallel to two unusual bright bands on Europa, Agenor Linea<sup>4,5</sup> and Corick Linea<sup>13</sup>. Both bands are also antipodal to each other and lie midway between each circle. This symmetry is unlikely to be mere coincidence. On either side of Agenor Linea (and the great circle) lie two zones of orthogonal strike-slip and pull-apart faulting associated with dark wedge-shaped bands<sup>5</sup>; this faulting involves formation of >100 km of new lithospheric material. Stratigraphically, Agenor Linea and the pull-apart bands are intermediate to young in age<sup>1,11</sup>. If SCDs, pull-apart bands and bright bands are related, then the SCDs might also date from this period.

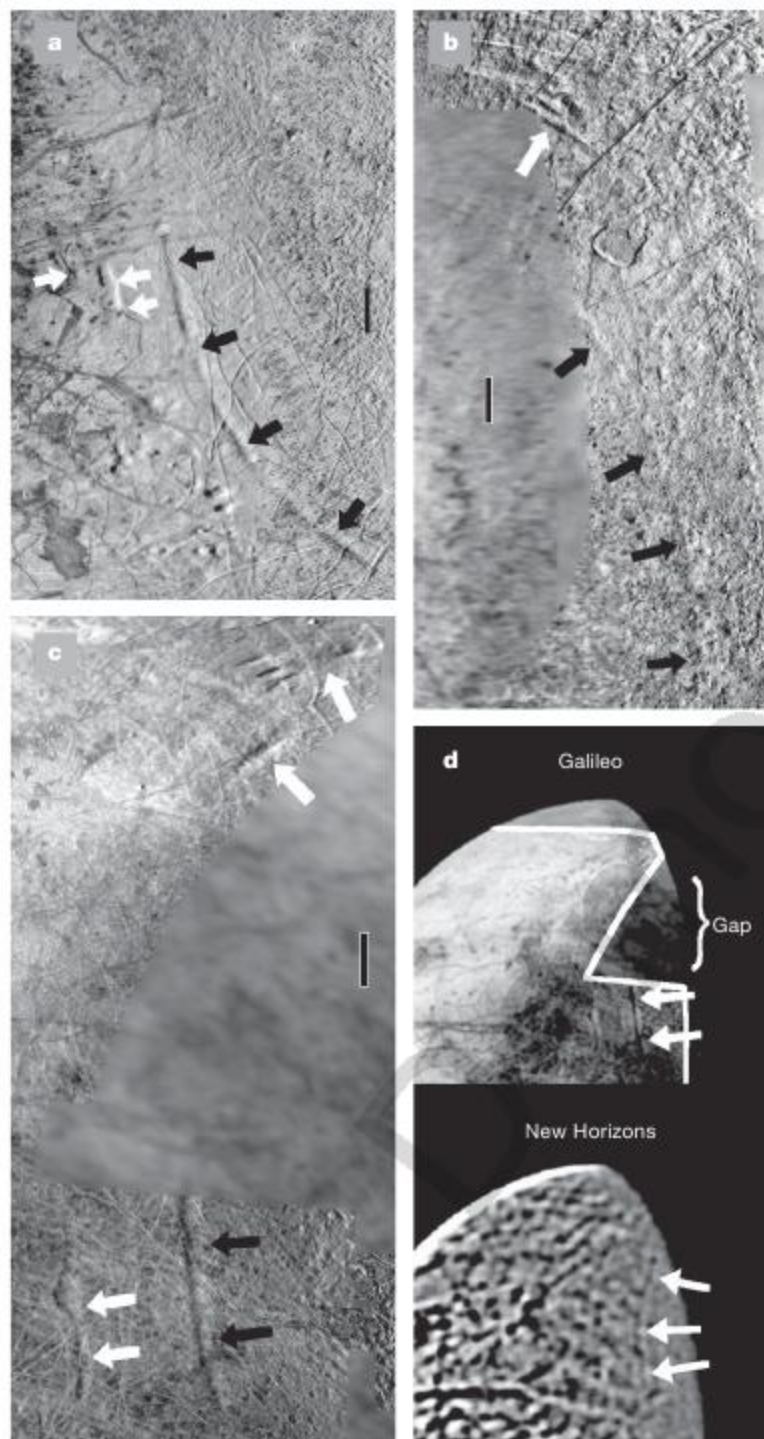
The global nature and symmetry of the SCDs suggests that global stresses are involved, most probably those due to global expansion/contraction<sup>14</sup>, tidal flexing<sup>15</sup>, despinning<sup>16</sup> and/or reorientation. Reorientation can include non-synchronous rotation<sup>17</sup> about the polar axis, and true polar wander<sup>2</sup> (TPW), involving rotation of the polar regions to the equator. Expansion/contraction gives rise to isotropic stress patterns<sup>14</sup>. The orientations of large-scale lineaments on Europa, specifically cycloid ridges and triple bands, are consistent with diurnal stresses<sup>18</sup> and ordinary non-synchronous rotation<sup>15,16,19</sup>, respectively, although we note that these global stress patterns (as well as despinning stresses) are symmetric about the equator, unlike our features. Accordingly, we consider that stresses due to reorientation of the surface relative to the principal axes<sup>2,17</sup> are the most likely cause of the circles.

Reorientations due to TPW and to non-synchronous rotation are both dynamically possible on Europa<sup>2,17</sup>. Non-synchronous rotation

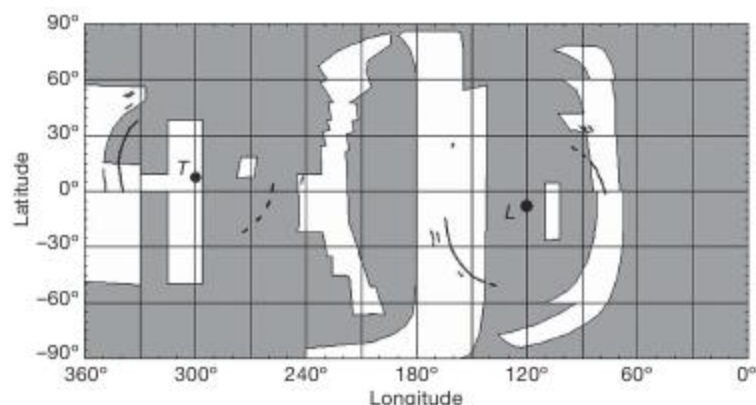
<sup>1</sup>Lunar and Planetary Institute, Houston, Texas 77058, USA. <sup>2</sup>Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA. <sup>3</sup>Department of Earth & Planetary Sciences, University of California, Santa Cruz, California 95064, USA.



generates compressional features centred on the initial tidal axes<sup>3,20</sup>, which makes the offset location and inferred extensional nature of the SCDs hard to explain. Furthermore, as the SCDs are not centred on or symmetric about the current equator, some TPW must have occurred, and TPW results in extensional stresses around the original rotation axis<sup>20,21</sup>. Pull-apart bands<sup>3</sup> and strike-slip fault and chaos distributions<sup>22</sup> may be related to TPW, but a definitive interpretation has not been possible because the limited Galileo imaging prevents global mapping of these smaller-scale features. Reorientations



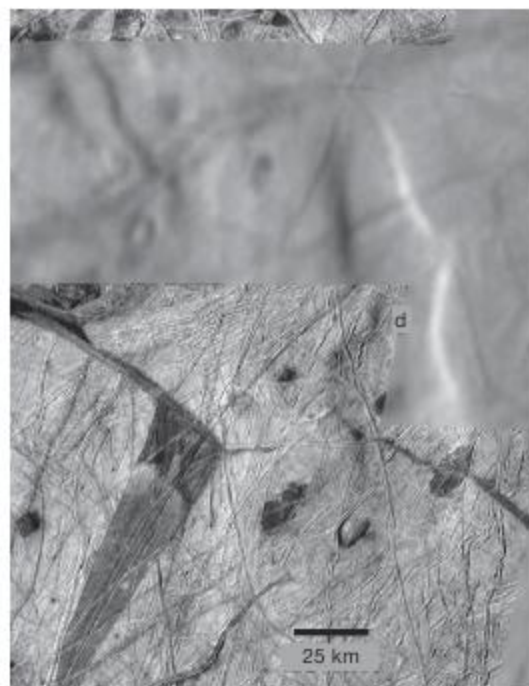
**Figure 1 | Regional views of global-scale arcuate troughs and basins on Europa.** **a**, Southwest quadrant, circle L (L, leading hemisphere: view centred at 25° S, 160° W); **b**, northeast quadrant, circle L (view centred at 12° N, 85° W); **c**, northwestern quadrant, circle T (T, trailing hemisphere: view centred at 27° N, 340° W). Main arcuate troughs (black arrows) and secondary basins (white arrows) are highlighted. Both types are ~20–50 km wide but secondary basins are only 100–300 km long. Vertical scale bars are 100 km in these cylindrical map projections. **d**, New Horizons imaging<sup>7</sup> (bottom) confirms the northward continuation of the main arcuate trough (arrows) within the Galileo mapping coverage gap (top, and blurry area of **c**). North is to top and the Sun is to the west (left) in all images.



**Figure 2 | A global map of the distribution of large-scale arcuate troughs and basins on Europa.** Black curves, locations of arcuate troughs and secondary basins. Low-Sun regional Voyager and Galileo imaging of the type required to observe these features (white regions) covers approximately 20–30% of the surface in widely dispersed north–south longitudinal bands at 220–1,500 m per pixel resolution. New Horizons<sup>7</sup> acquired imaging at 17–20 km resolution between longitudes 270° and 360° W, showing a northern extension of the trough segment at 340° longitude into the Voyager-Galileo data gap. Dots are least-squares centres of curvature for circle L (L) and circle T (T). Dashed curve at 270° W is the predicted location of the eastern arcuate trough segment of circle T, unobserved by Galileo and not detected by New Horizons. Longitude is in degrees west; positive and negative latitude denotes degrees north and south, respectively.

approaching  $90^\circ$  have been predicted for Europa<sup>3</sup> and are more likely to occur around the tidal axis<sup>23</sup>. Stresses developed during TPW are typically an order of magnitude (or more) larger than diurnal tidal stresses<sup>2</sup>. Large TPW, with the initial rotation axis passing near the meridians containing the centres of the SCDs, is required to explain the location of the SCDs.

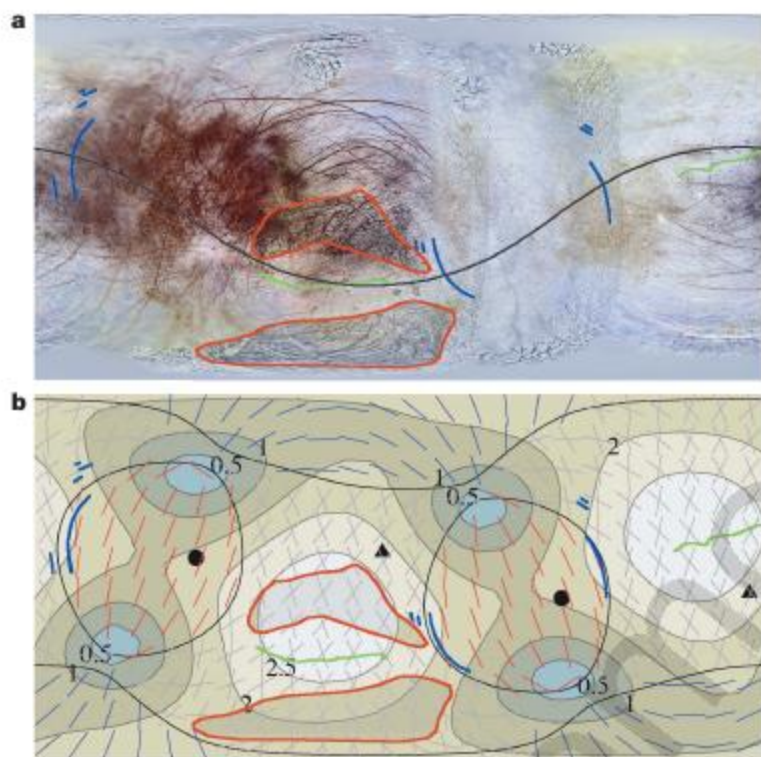
Figure 4 plots the stress contours and fracture patterns arising from 80° of TPW roughly around the tidal axis, combined with a small amount (0.002%) of isotropic expansion (Supplementary



**Figure 3 | Highest-resolution image mosaic of a large secondary basin on Europa.** Shown is a high resolution (230 m per pixel) Galileo image mosaic of the southern part of a large secondary basin (d) in the southwest quadrant of circle L. Low-resolution data in the upper section is global context imaging at 1.5 km per pixel. Continuity of ridges and dark features across the margin of the basin and a lack of unusual resurfacing on its floor are evident. Scene is an enlargement of the centre left portion of Fig. 1a. Image resolution (230 m per pixel) is not sufficient to resolve the nature of the steep margins, which most probably are due to unresolved normal faults or to monoclinical folding.



Information). We arrived at a solution with the palaeopole at  $10^\circ$  N,  $280^\circ$  W, and a self-consistent reorientation of the tidal axis such that the initial sub-Jupiter point ( $0^\circ$ ,  $0^\circ$  W) is now at  $10^\circ$  S,  $8.2^\circ$  W. Although there are similar solutions that can match the size and location of the SCDs, all of them involve a large TPW and a small, self-consistent reorientation of the tidal axis. We chose the solution with the palaeopole at  $10^\circ$  N,  $280^\circ$  W because, in addition to matching the location and size of the SCDs, the deviatoric stress is maximum where depressed segments of the circles are observed. If the arcuate troughs are extensional, then they are likely to form first within the regions dominated by extensional faulting where the maximum extensional stresses occur. The locations and orientations of the SCDs follow this prediction very closely (Fig. 4). If Agenor and Corick Lineae are strike-slip duplex features<sup>4</sup>, their locations (and perhaps their orientations) are also consistent with the predicted stresses, as are the locations and orientations of the orthogonal pull-apart bands (Fig. 4b).



**Figure 4 | Global colour mosaic and predicted stress patterns for true polar wander in a thin elastic shell on Europa.** **a**, Colour mosaic. Blue curves are SCDs, wavy green lines are bright curvilinear bands Agenor Linea ( $180\text{--}240^\circ$  W longitude) and Corick Linea ( $0\text{--}50^\circ$  W longitude); red outlined regions are pull-apart band dilatational and strike-slip zones<sup>5</sup>. A great circle (heavy black curve) roughly bisecting the long arcuate troughs and the two strike-slip zones is also shown. Long arcuate triple bands lie either side of these zones and converge towards the centres of the two circles. **b**, Stress/fracture map. Palaeopole locations (solid dots: palaeo-north-pole  $10^\circ$  N,  $280^\circ$  W) and initial tidal axes (triangles: palaeo-sub-Jupiter-point  $10^\circ$  S,  $8.2^\circ$  W) are shown. Total angular reorientation is  $80^\circ$ ; isotropic extensional strain is  $0.002\%$ . The short red lines, grey crosses and blue lines indicate directions of the expected normal, strike-slip and thrust faulting, respectively. Black lines separate these tectonic regions. Two zones of extensional stress are separated by  $180^\circ$  and offset in latitude by  $20^\circ$ , separated by regions in which strike-slip motions dominate. The centres of the extensional regions are offset in longitude from the meridians containing the initial rotation axis because the reorientation geometry requires a small ( $8.2^\circ$ ) eastward longitudinal reorientation of the tidal axis such that the rotational and tidal axes remain perpendicular. Pull-apart bands are associated with both dilation and strike-slip faulting oriented northeast–southwest and northwest–southeast<sup>5</sup>, and Agenor Linea may be a strike-slip zone in combination with extension or compression<sup>4</sup>, both consistent with the predicted pattern. Shaded regions and contours correspond to deviatoric stress in units of MPa. Stresses were calculated assuming a shear modulus of  $9.3$  GPa (ref. 22) and Love number  $h = 1$ . We note that the fit between feature and stress patterns may be enhanced if a further TPW shell rotation of  $\sim 15^\circ$  is assumed.

Assuming that the secondary basins are fault-bounded, a maximum depth of  $1$  km implies a total extension of  $\sim 2$  km, or a global strain of  $\sim 0.02\%$ . An isostatically supported trough  $1$  km deep would imply a minimum shell thickness of  $6\text{--}17$  km (ref. 8), depending on shell and ocean composition, although short shell-flow timescales<sup>24</sup> suggest that flexural support may be a more likely mechanism.

The SCDs form a striking set of features unique to Europa. An  $\sim 80^\circ$  reorientation event involving TPW explains the observed locations, orientations and characteristics of the SCDs. Large-scale lineaments such as pull-apart bands, bright bands, as well as triple bands (currently thought to have formed during ordinary non-synchronous rotation<sup>19</sup>), form a symmetric pattern with respect to TPW stress patterns, and can also potentially be explained as extensional or strike-slip features formed during this TPW episode. These correlations suggest that the origins of many of Europa's tectonic and diapiric features<sup>1,11</sup> should be re-evaluated to determine whether they also formed, or were displaced<sup>22</sup>, during the reorientation proposed here. This can partly be accomplished with Galileo's severely limited mapping coverage, but will ultimately require a new high-resolution global mapping mission to Europa.

Polar reorientation by  $\sim 70\text{--}90^\circ$  is an expected consequence of latitudinally varying thickness in a decoupled ice shell on Europa<sup>2,3</sup>, although dissipation within the ice shell is likely to reduce this reorientation somewhat<sup>25</sup>. TPW on Europa also provides an independent line of evidence that the ice shell is (or once was) floating on a sub-surface water ocean<sup>2</sup>. Our hypothesized episode of TPW places Europa in the company of Mars<sup>26,27</sup>, Enceladus<sup>28</sup> and perhaps Miranda<sup>29</sup>, and may have fundamentally reshaped Europa's geology.

Received 18 July 2007; accepted 6 March 2008.

- Greeley, R. et al. in *Jupiter* (eds Bagenal, F., Dowling, T. & McKinnon, W.) 329–362 (Cambridge Univ. Press, Cambridge, UK, 2004).
- Ojakangas, G. & Stevenson, D. Polar wander of an ice shell on Europa. *Icarus* **81**, 242–270 (1989).
- Leith, A. & McKinnon, W. Is there evidence for polar wander on Europa? *Icarus* **120**, 387–398 (1996).
- Prockter, L., Pappalardo, R. & Head, J. Strike-slip duplexing on Jupiter's icy moon Europa. *J. Geophys. Res.* **105**, 9483–9488 (2000).
- Schenk, P. & McKinnon, W. Fault offsets and lateral crustal movement on Europa: Evidence for a mobile ice shell. *Icarus* **79**, 75–100 (1989).
- Schenk, P. & McKinnon, W. Ring geometry on Ganymede and Callisto. *Icarus* **72**, 209–234 (1987).
- Grundy, W. et al. New Horizons mapping of Europa and Ganymede. *Science* **318**, 234–238 (2007).
- Schenk, P. & McKinnon, W. Topographic variability on Europa from Galileo stereo images. *Lunar Planet. Sci. Conf.* **32**, abstr. 2078 (2001).
- Figueredo, P. & Greeley, R. Geologic mapping of the northern leading hemisphere of Europa from Galileo solid-state imaging data. *J. Geophys. Res.* **105**, 22629–22646 (2000).
- Prockter, L. & Schenk, P. Origin and evolution of Castalia Macula, an anomalous young depression on Europa. *Icarus* **177**, 305–326 (2005).
- Figueredo, P. & Greeley, R. Resurfacing history of Europa from pole-to-pole geologic mapping. *Icarus* **167**, 287–312 (2004).
- Schenk, P. Crop circles of Europa. *Lunar Planet. Sci. Conf.* **30**, abstr. 2081 (2005).
- Greenberg, R. The evil twin of Agenor: Tectonic convergence on Europa. *Icarus* **167**, 313–319 (2004).
- Nimmo, F. Stresses generated in cooling viscoelastic ice shells: Application to Europa. *J. Geophys. Res.* **109**, E12001, doi:10.1029/2004JE002347 (2004).
- Greenberg, R. et al. Tectonic processes on Europa: Tidal stresses, mechanical response, and visible features. *Icarus* **135**, 64–78 (1998).
- Melosh, H. J. Global tectonics of a despun planet. *Icarus* **31**, 221–243 (1977).
- Greenberg, R. & Weidenschilling, S. How fast do Galilean satellites spin? *Icarus* **58**, 186–196 (1984).
- Hoppa, G., Tufts, B. R., Greenberg, R. & Geissler, P. Formation of cycloidal features on Europa. *Science* **285**, 1899–1902 (1999).
- Geissler, P. et al. Evidence for non-synchronous rotation of Europa. *Nature* **391**, 368–370 (1998).
- Matsuyama, I. & Nimmo, F. Tectonic patterns of a reoriented and despun planetary bodies. *Icarus* doi:10.1016/j.icarus.2007.12.003 (in the press).
- Melosh, H. J. Tectonic patterns on a reoriented planet — Mars. *Icarus* **44**, 745–751 (1980).



22. Sarid, M. et al. Polar wander and surface convergence of Europa's ice shell: Evidence from a survey of strike-slip displacement. *Icarus* **158**, 24–41 (2002).
23. Matsuyama, I. & Nimmo, F. Rotational stability of tidally deformed planetary bodies. *J. Geophys. Res.* **112**, E11003, doi:10.1029/2007JE002942 (2007).
24. Stevenson, D. Limits on the variation of thickness of Europa's ice shell. *Lunar Planet. Sci. Conf.* **31**, abstr. 1506 (2000).
25. Matsuyama, I., Nimmo, F. & Mitrovica, J. Reorientation of planets with lithospheres: The effect of elastic energy. *Icarus* **191**, 401–412 (2007).
26. Willemann, R. Reorientation of planets with elastic lithospheres. *Icarus* **60**, 701–709 (1984).
27. Matsuyama, I., Mitrovica, J., Manga, M., Perron, J. & Richards, M. Rotational stability of dynamic planets with elastic lithospheres. *J. Geophys. Res.* **111**, E02003, doi:10.1029/2005JE002447 (2006).
28. Nimmo, F. & Pappalardo, R. Diapir-induced reorientation of Saturn's moon Enceladus. *Nature* **441**, 614–616 (2006).
29. Janes, D. M. & Melosh, H. J. Sinkers tectonics — an approach to the surface of Miranda. *J. Geophys. Res.* **93**, 3127–3143 (1988).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported in part by the NASA Outer Planets Research and Planetary Geology and Geophysics programmes. LPI contribution no. 1399.

**Author Contributions** Global mapping and topographic analyses were conducted by P.S.; P.S. and F.N. conducted preliminary examination of alternative models; numerical modelling of stress patterns was conducted by I.M.; and P.S. wrote the paper, with contributions from I.M. and F.N.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.S. ([schenk@lpi.usra.edu](mailto:schenk@lpi.usra.edu)).

Demo mode



## LETTERS

# A GaAs polariton light-emitting diode operating near room temperature

S. I. Tsintzos<sup>1</sup>, N. T. Pelekanos<sup>1,3</sup>, G. Konstantinidis<sup>3</sup>, Z. Hatzopoulos<sup>2,3</sup> & P. G. Savvidis<sup>1,3</sup>

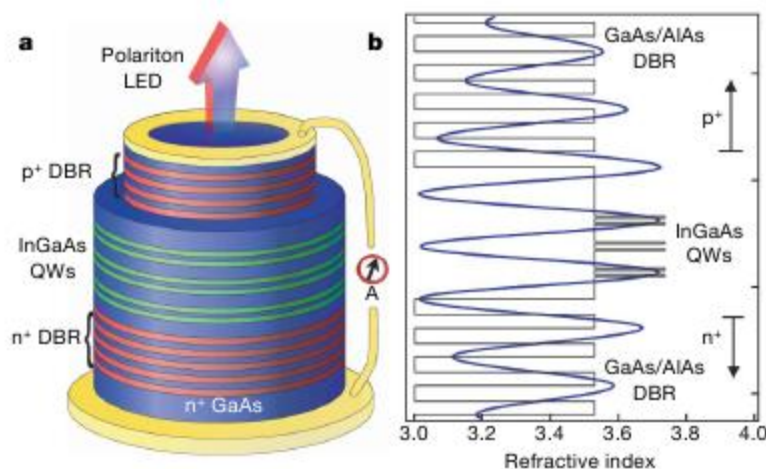
The increasing ability to control light-matter interactions at the nanometre scale has improved the performance of semiconductor lasers in the past decade. The ultimate optimization is realized in semiconductor microcavities, in which strong coupling between quantum-well excitons and cavity photons gives rise to hybrid half-light/half-matter polariton quasiparticles<sup>1</sup>. The unique properties of polaritons—such as stimulated scattering<sup>2,3</sup>, parametric amplification<sup>4–6</sup>, lasing<sup>7–10</sup>, condensation<sup>11–13</sup> and superfluidity<sup>14,15</sup>—are believed to provide the basis for a new generation of polariton emitters and semiconductor lasers. Until now, polariton lasing and nonlinearities have only been demonstrated in optical experiments, which have shown the potential to reduce lasing thresholds by two orders of magnitude compared to conventional semiconductor lasers<sup>16</sup>. Here we report an experimental realization of an electrically pumped semiconductor polariton light-emitting device, which emits directly from polariton states at a temperature of 235 K. Polariton electroluminescence data reveal characteristic anticrossing between exciton and cavity modes, a clear signature of the strong coupling regime. These findings represent a substantial step towards the realization of ultra-efficient polaritonic devices with unprecedented characteristics.

The study of polaritons in semiconductor microcavities has led to a wealth of experimental demonstrations in the past ten years, including polariton stimulation<sup>2,3</sup>, amplification<sup>4–6</sup> and lasing<sup>7–10</sup>. In a polariton laser, the emission of coherent monochromatic light requires no conventional population inversion scheme, owing to the bosonic nature of exciton-polaritons, which enables them to condense in a single quantum state<sup>17</sup>. The coherence in the emitted light is inherited directly from this macroscopically coherent state through radiative decay of polaritons. However, until now polariton lasing and nonlinearities have only been demonstrated in optical experiments (using either non-resonant excitation or an efficient stimulated polariton scattering mechanism to generate polaritons), which show the potential of polariton lasers for lasing thresholds reduced by two orders of magnitude compared to conventional semiconductor lasers<sup>16</sup>. In the context of light sources, polariton devices based on electrical injection, rather than optical pumping, would be more suitable for real-world applications. A number of key issues, such as the resistivity of p-type distributed Bragg reflector (DBR) mirrors at low temperatures<sup>18</sup>, cavity losses related to doping of the mirrors, and polariton stability under electrical injection, have thwarted so far the realization of any electrically pumped polariton emitting semiconductor device.

Electrical injection of a microcavity light-emitting diode (LED) in the strong coupling regime has only been reported recently in organic semiconductors<sup>19</sup>. The drawback of organic microcavities is that the electroluminescence spectra are superposed on emission from the

hole-injection diamine derivative (TPD) layer and localized excitonic states that do not strongly couple to cavity photons, restricting severely the purity and efficiency of polaritonic emission. Furthermore, a different type of intersubband polariton has been reported recently in quantum cascade structures<sup>20</sup> along with a demonstration of a mid-infrared polariton quantum cascade LED device<sup>21</sup>. However, the validity of the bosonic description for intersubband polaritons remains to be tested experimentally, raising fundamental questions about whether it is possible to produce a polariton-based laser.

The microcavity LED structure of this work was grown by molecular beam epitaxy on a GaAs (001)  $n^+$  substrate (Fig. 1a). The bottom DBR mirror consists of 21 periods of GaAs/AlAs  $n$ -type doped to  $2 \times 10^{18} \text{ cm}^{-3}$ , whereas the top DBR consists of 17 periods of GaAs/AlAs  $p$ -type doped to  $4 \times 10^{18} \text{ cm}^{-3}$ . The  $5\lambda/2$  GaAs cavity contains three pairs of  $\text{In}_{0.1}\text{Ga}_{0.9}\text{As}/\text{GaAs}$  quantum wells placed at the antinodes of the electric field (Fig. 1b). Polariton LED mesas with 400  $\mu\text{m}$  diameter were dry etched using  $\text{BCl}_3/\text{Cl}_2$  reactive ion etching. The etching was deep enough to remove the top  $p$ -type DBR and ensure electrical isolation of the device. An  $n$ -type (Au/Ge) ohmic contact was evaporated on the back side of the substrate and a ring-shaped  $p$ -type (Ti/Pt) ohmic contact was deposited on top of the mesa. The

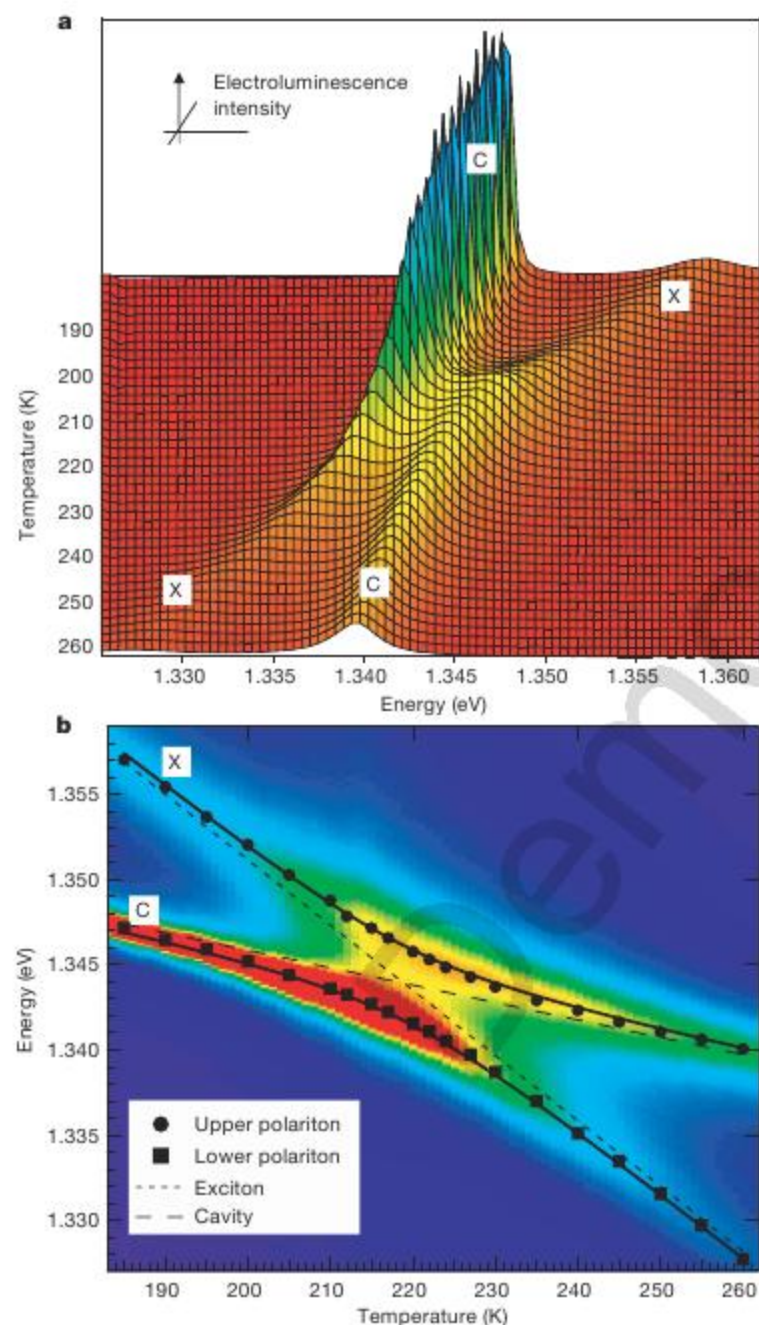


**Figure 1 | Schematic sketch of the polariton microcavity LED.** **a**, The structure consists of a  $5\lambda/2$  cavity surrounded by two doped GaAs/AlAs DBRs. Dry etching was used to expose the  $p$ -DBR, forming round mesas of 400  $\mu\text{m}$  diameter. To enable light emission from the front side, a ring-shaped  $p$ -type ohmic contact was deposited on the top of the mesa, and the  $n$ -type contact was deposited on the back side of the substrate. QW, quantum well. Current source  $A$  is used to bias the device. **b**, Electric field and refractive index distribution along the structure. Elevated temperatures necessitate the use of three pairs of  $\text{In}_{0.1}\text{Ga}_{0.9}\text{As}/\text{GaAs}$  quantum wells placed at the antinodes of the electric field to enhance the exciton oscillator strength required for strong coupling.

<sup>1</sup>Department of Materials Science & Technology, <sup>2</sup>Department of Physics, University of Crete, PO Box 2208, 71003 Heraklion, Crete, Greece. <sup>3</sup>Foundation for Research and Technology-Hellas, Institute of Electronic Structure & Laser, PO Box 1527, 71110 Heraklion, Crete, Greece.



sample was designed to exhibit the strong coupling regime at elevated temperatures to avoid injection problems associated with the high resistivity of p-type DBRs at low temperatures. To minimize the doping-related losses inside the cavity, the first half period of the DBRs, where penetration of the electric field into the mirror is considerable, was kept undoped. Reflectivity measurements on the sample show two coherently coupled exciton-photon states separated by a Rabi energy of 4.4 meV at a temperature of 220 K, demonstrating that doping-induced cavity mode broadening is not a restricting factor in our structure. To resolve the rich angular properties of our polariton emitter, a specially designed goniometer was built to collect light from a wide range of angles. Angle-resolved electroluminescence measurements were performed in a temperature-controlled, closed-cycle, electrically wired liquid helium cryostat.



**Figure 2 | Polariton electroluminescence emission versus temperature.** **a**, Electroluminescence spectra from the polariton LED structure for different values of temperature at zero angle and an injection current of 0.4 mA. Relative tuning of the exciton resonance X through the cavity mode C is achieved by exploiting different energy shifts of the two modes with temperature. Characteristic anticrossing behaviour is revealed. **b**, Contour plot and extracted peaks of the polariton emission shown in **a**, resolving the two polariton branches as a function of temperature. The solid lines are theoretical fits, while the dashed and dotted lines are the bare exciton and cavity modes, respectively. Rabi splitting of 4.4 meV is observed at 219 K.

The light was collected through an optical fibre mounted on one of the goniometer arms.

The demonstration of the strong coupling regime in a variety of semiconductor systems via measurements of reflectivity, transmission, or photo- or electroluminescence involves the ability to tune either of the excitonic and cavity mode resonances with respect to the other, showing the characteristic anticrossing/coupling between the two modes. To study emission properties of our microcavity polariton LED device under electrical injection, we used both temperature and angular tuning measurements to prove that emission arises truly from polariton states. The emission energy of the quantum well excitons follows the temperature dependence of the GaAs bandgap (about  $-0.38 \text{ meV K}^{-1}$  at 220 K), whereas the corresponding cavity mode energy shift is due to the change of the refractive index and is much weaker (about  $-0.102 \text{ meV K}^{-1}$  at 220 K).

Figure 2a shows electroluminescence spectra from the polariton LED for different temperatures between 180 and 260 K, collected at zero angle. The emission at 180 K consists of a high energy quantum well exciton line X at 1.357 eV and a low energy cavity mode line C at 1.347 eV. The increase of the sample's temperature by 80 K tunes the exciton energy by 29 meV down to 1.328 eV, and the cavity mode by 7.4 meV down to 1.34 eV. Notably, at 260 K the low energy line has transformed into excitonic emission whereas the high energy line has transformed into cavity mode emission. The exciton and cavity modes are well resolved over the entire temperature range, and exhibit the characteristic anticrossing behaviour of the strong coupling regime. This shows that electroluminescence in this intermediate temperature range arises unambiguously from exciton polariton states in the strong coupling regime, which is demonstrated for the first time in electrically pumped semiconductor microcavities. It should be noted that although excitonic transitions in bare  $\text{In}_{0.1}\text{Ga}_{0.9}\text{As}/\text{GaAs}$  quantum wells are resolved in absorption measurements up to room temperature, the emission from these structures is largely dominated at elevated temperatures by free electron hole pair recombination. These findings are remarkable, and suggest that by dressing the exciton with a photon in a semiconductor microcavity, the excitonic contribution to the emission becomes dominant even at high temperatures.

To shed more light on the dispersive features of the two polariton branches, the peak positions from the electroluminescence data are extracted and plotted in Fig. 2b. An excellent fit of the upper and lower polariton branches is obtained by applying a coupled harmonic oscillator model to our data. In this model, the temperature dependence of the exciton and cavity modes has been taken into account, as described above. Small deviations from experiment could be well understood by considering that no dependence of exciton oscillator strength on temperature has been included in the model. The fits reveal that the condition of zero exciton-photon detuning at zero angle is reached at a temperature of 219 K with a normal mode Rabi splitting of 4.4 meV, in agreement with the reflectivity spectra taken on the same mesa at the same temperature.

To investigate polariton relaxation dynamics under electrical injection at this elevated temperature, electroluminescence is recorded as a function of angle for a fixed current of  $I = 0.2 \text{ mA}$ , at temperatures of 220 K and 195 K, achieving zero and negative ( $\delta = -7 \text{ meV}$ ) detuning conditions, respectively. Selected characteristic spectra at different angles are plotted in Fig. 3a, b. The extracted peak positions, resolving the complete polariton dispersions at these different detunings, are shown in Fig. 3c, d. For both detuning conditions, clear anticrossing between the two polariton branches (upper and lower) is resolved. In Fig. 3e, f, polar plots of the integrated lower polariton branch emission for two different injection currents are shown. For zero detuning, a ring-shaped emission at  $15^\circ$  is observed at low injection current, indicating the presence of a bottleneck effect, often found in similar structures at low temperatures, due to the suppressed relaxation of exciton polaritons on the lower polariton branch<sup>22</sup>. The relaxation bottleneck at zero detuning appears to be at



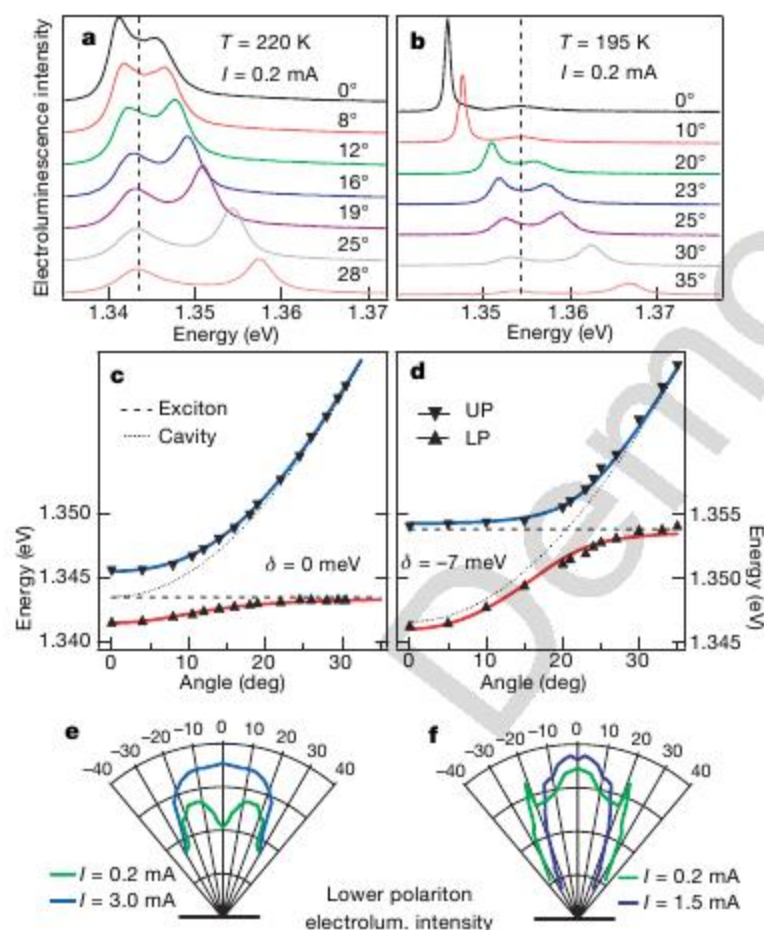
least partially suppressed when increasing injection current to 3.0 mA, due to increased inter-carrier scattering, which leads to faster relaxation of polaritons along the branch, enhancing the emission into the normal direction. More pronounced bottleneck suppression is found for negative detuning, where the ring-shaped emission peaked at  $21^\circ$  at low current is completely suppressed with increased injection. Bypassing relaxation bottleneck effects is a crucial step towards the development of an electrically pumped polariton laser.

A legitimate question to ask is whether increasing injection density will lead to polariton lasing or to a collapse of the strong coupling regime and transition to conventional vertical cavity surface-emitting laser (VCSEL) operation. To achieve the higher injection densities needed for such a test, we resort to a neighbouring device on the same wafer with shifted cavity mode resonance, which exhibits the zero-detuning condition in the normal direction at 235 K. The injection is improved at higher operation temperatures because of the lower series resistance of the p-type DBR. In Fig. 4a, electroluminescence spectra normalized by current are shown as a function of injection current. At low injection currents, two clearly resolved polariton peaks are evident. With increasing injection current, the two polariton peaks progressively merge into a single emission line, underscoring the transition to the weak coupling regime (Fig. 4b), that is, that of a VCSEL operation. The integrated electroluminescence emission from both polariton peaks shows a superlinear (quadratic) increase (Fig. 4c). To see if this superlinearity is of an

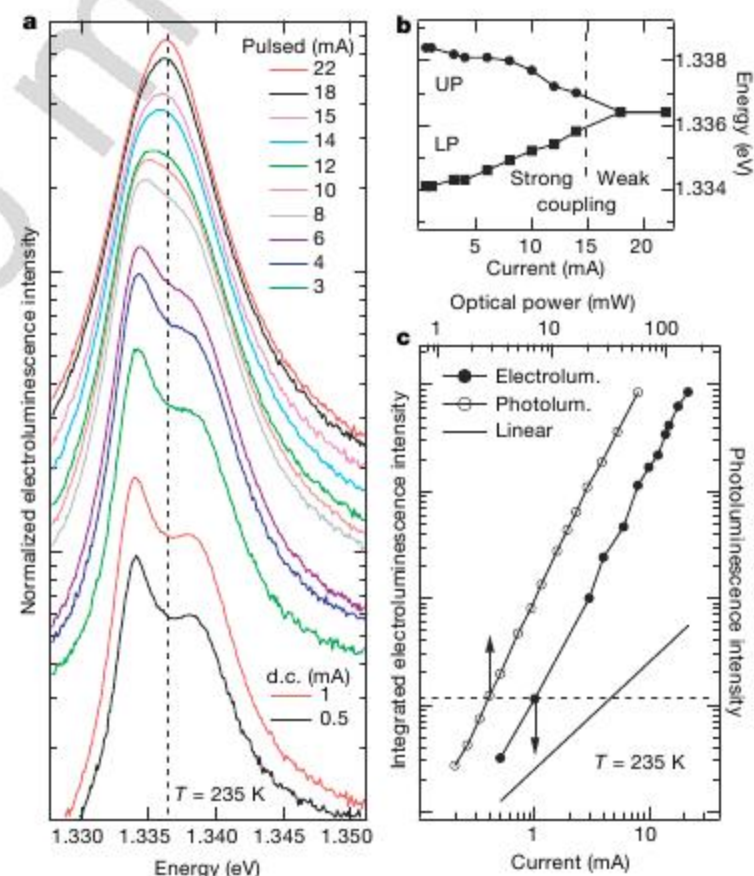
intrinsic character, we performed optical non-resonant excitation experiments under the same conditions using a 846 nm continuous-wave laser. Identical superlinearity is observed in the optical experiments, as shown by circles in Fig. 4c, strongly suggesting that its origin is of an intrinsic character. The observed superlinear dependence on current could be attributed to the enhanced relaxation of polaritons down the lower polariton branch with increasing injection, as pointed out in the discussion of Fig. 3.

By measuring the optical power for which the photoluminescence emission intensity is identical to the electroluminescence intensity at  $I = 1$  mA (Fig. 4c), we estimate that at the maximum current injection of 22 mA in our experiments an estimated density of  $10^{10}$  polaritons  $\text{cm}^{-2}$  are injected per quantum well. This density compares well to the exciton saturation density of  $3 \times 10^{10}$  polaritons  $\text{cm}^{-2}$  per quantum well<sup>14</sup>, considering the much higher temperatures in our experiments. The above observations indicate that the present injection scheme is unlikely to yield a polariton laser, owing to inefficient polariton relaxation down the polariton branch. A possible approach would be to bypass the relaxation bottleneck, possibly by using fast longitudinal optical phonon enhanced relaxation in a different device geometry.

Our findings open the way for the development of a new generation of polariton emitters and novel forms of lasing in semiconductors, using the well-established and mature GaAs technology. The advantages of electrically pumped polariton emitting devices operating in the strong coupling regime are many. First,



**Figure 3 | Angle-dependent polariton electroluminescence.** **a, b**, Sets of selected spectra under electrical pumping as a function of collection angle for zero and negative ( $\delta = -7$  meV) detuning conditions, respectively.  $T$ , temperature;  $I$ , current. **c, d**, Polariton energy dispersions. Excellent fits (solid lines) are obtained using a coupled harmonic oscillator model for different detunings. Bare exciton and cavity mode energies are shown dashed. UP, upper polariton; LP, lower polariton. **e, f**, Polar plots of the integrated lower polariton emission at various angles obtained at various currents. Ring-shaped emission at  $15^\circ$  ( $21^\circ$ ) is observed at low injection for zero (negative) detuning, indicating the presence of a bottleneck. Partial suppression of the relaxation bottleneck is observed at higher injection currents for both detunings.



**Figure 4 | Current-dependent polariton electroluminescence.** **a**, Normalized electroluminescence spectra with injection currents shown, obtained in the normal direction at 235 K for zero detuning. At higher currents, a pulsed current source with a duty cycle of 1:200 is used to eliminate heating effects. **b**, Extracted lower and upper polariton peak positions, showing the onset of the weak coupling regime at high currents. **c**, Integrated electroluminescence intensity of both the lower and upper polariton peaks (filled symbols), showing a quadratic superlinear dependence. For comparison, the solid line represents a linear dependence. In open symbols are shown the integrated photoluminescence peak intensities obtained for various optical powers of excitation, showing the same superlinear dependence.



they offer dramatic enhancement of spontaneous emission rate even when compared to conventional resonant cavity light-emitting diodes (RCLEDs). In the strong coupling regime, the characteristic time in light–matter interactions is set by Rabi oscillation rather than the radiative lifetime. This makes polariton-based emitters ideal candidates to become future LED devices of unprecedented quantum efficiency and ultralow power consumption. Second, the four orders of magnitude lower density of states of polaritons allows stimulation to be achieved at much lower carrier densities compared to other semiconductor lasers. It is expected that polariton lasers will have at least one order of magnitude lower lasing thresholds. The essential element in all the above devices is the ability to electrically pump polaritons without destroying the strong coupling regime at elevated temperatures—as demonstrated here. We note that after this manuscript was submitted, we became aware of recent reports of polariton electroluminescence in GaAs microcavities<sup>23,24</sup>, which, however, is restricted to low temperatures.

Received 23 December 2007; accepted 8 April 2008.

1. Weisbuch, C., Nishioka, M., Ishikawa, A. & Arakawa, Y. Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. *Phys. Rev. Lett.* **69**, 3314–3317 (1992).
2. Dang, L. S., Heger, D., André, R., Bœuf, F. & Romestain, R. Stimulation of polariton photoluminescence in semiconductor microcavity. *Phys. Rev. Lett.* **81**, 3920–3923 (1998).
3. Senellart, P., Bloch, J., Sermage, B. & Marzin, J. Y. Microcavity polariton depopulation as evidence for stimulated scattering. *Phys. Rev. B* **62**, R16263–R16266 (2000).
4. Savvidis, P. G. *et al.* Angle-resonant stimulated polariton amplifier. *Phys. Rev. Lett.* **84**, 1547–1550 (2000).
5. Ciuti, C., Schwendimann, P., Deveaud, B. & Quattropani, A. Theory of the angle-resonant polariton amplifier. *Phys. Rev. B* **62**, R4825–R4828 (2000).
6. Saba, M. *et al.* High-temperature ultrafast polariton parametric amplification in semiconductor microcavities. *Nature* **414**, 731–735 (2001).
7. Porras, D., Ciuti, C., Baumberg, J. J. & Tejedor, C. Polariton dynamics and Bose-Einstein condensation in semiconductor microcavities. *Phys. Rev. B* **66**, 085304 (2002).
8. Malpuech, G., Di Carlo, A. & Kavokin, A. Room-temperature polariton lasers based GaN microcavities. *Appl. Phys. Lett.* **81**, 412–414 (2002).
9. Kasprzak, J. *et al.* Bose-Einstein condensation of exciton polaritons. *Nature* **443**, 409–414 (2006).
10. Christopoulos, S. *et al.* Room-temperature polariton lasing in semiconductor microcavities. *Phys. Rev. Lett.* **98**, 126405 (2007).
11. Deng, H. *et al.* Quantum degenerate exciton-polaritons in thermal equilibrium. *Phys. Rev. Lett.* **97**, 146402 (2006).
12. Butov, L. A polariton laser. *Nature* **447**, 540–541 (2007).
13. Balili, R., Hartwell, V., Snoke, D., Pfeiffer, L. & West, K. Bose-Einstein condensation of microcavity polaritons in a trap. *Science* **316**, 1007–1010 (2007).
14. Carusotto, I. & Ciuti, C. Probing microcavity polariton superfluidity through resonant Rayleigh scattering. *Phys. Rev. Lett.* **93**, 166401 (2004).
15. Kavokin, A., Malpuech, G. & Laussy, F. P. Polariton laser and polariton superfluidity in microcavities. *Phys. Lett. A* **306**, 187–199 (2003).
16. Deng, H., Weihs, G., Snoke, D., Bloch, J. & Yamamoto, Y. Polariton lasing vs. photon lasing in a semiconductor microcavity. *Proc. Natl Acad. Sci. USA* **100**, 15318–15323 (2003).
17. Imamoglu, A., Ram, R. J., Pau, S. & Yamamoto, Y. Nonequilibrium condensates and lasers without inversion: Exciton-polariton lasers. *Phys. Rev. A* **53**, 4250–4253 (1996).
18. Carlini, J. F. *et al.* Design and characterization of top-emitting microcavity light-emitting diodes. *Semicond. Sci. Technol.* **15**, 145–154 (2000).
19. Tischler, J. R., Bradley, M. S., Bulović, V., Song, J. H. & Nurmikko, A. Strong coupling in a microcavity LED. *Phys. Rev. Lett.* **95**, 036401 (2005).
20. Sapienza, L. *et al.* Photovoltaic probe of cavity polaritons in a quantum cascade structure. *Appl. Phys. Lett.* **90**, 201101 (2007).
21. Sapienza, L. *et al.* Electrically injected cavity polaritons. *Phys. Rev. Lett.* **100**, 136806 (2008).
22. Tartakovskii, A. I. *et al.* Relaxation bottleneck and its suppression in semiconductor microcavities. *Phys. Rev. B* **62**, R2283–R2286 (2000).
23. Bajoni, D. *et al.* Polariton light-emitting diode in a GaAs-based microcavity. *Phys. Rev. B* **77**, 113303 (2008).
24. Khalifa, A. A., Love, A. P. D., Krizhanovskii, D. N., Skolnick, M. S. & Roberts, J. S. Electroluminescence emission from polariton states in GaAs-based semiconductor microcavities. *Appl. Phys. Lett.* **92**, 061107 (2008).

**Acknowledgements** Support by the PENED projects 03EΔ841 and 03EΔ816 (funded 25% by national funds and 75% by EC) acknowledged.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.G.S. (psav@materials.uoc.gr).



## LETTERS

# Superconductivity at 43 K in an iron-based layered compound $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$

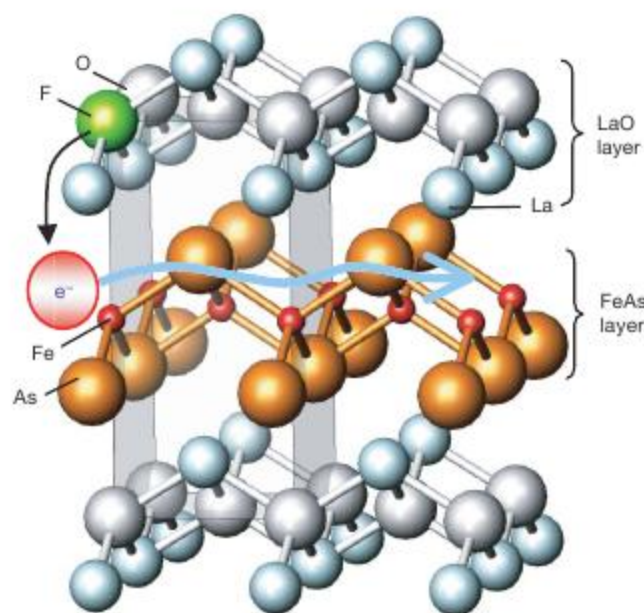
Hiroki Takahashi<sup>1</sup>, Kazumi Igawa<sup>1</sup>, Kazunobu Arii<sup>1</sup>, Yoichi Kamihara<sup>2</sup>, Masahiro Hirano<sup>2,3</sup> & Hideo Hosono<sup>2,3</sup>

The iron- and nickel-based layered compounds  $\text{LaOFeP}$  (refs 1, 2) and  $\text{LaONiP}$  (ref. 3) have recently been reported to exhibit low-temperature superconducting phases with transition temperatures  $T_c$  of 3 and 5 K, respectively. Furthermore, a large increase in the midpoint  $T_c$  of up to  $\sim 26$  K has been realized<sup>4</sup> in the isocrystalline compound  $\text{LaOFeAs}$  on doping of fluoride ions at the  $\text{O}^{2-}$  sites ( $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ ). Experimental observations<sup>5,6</sup> and theoretical studies<sup>7–9</sup> suggest that these transitions are related to a magnetic instability, as is the case for most superconductors based on transition metals. In the copper-based high-temperature superconductors, as well as in  $\text{LaOFeAs}$ , an increase in  $T_c$  is often observed as a result of carrier doping in the two-dimensional electronic structure through ion substitution in the surrounding insulating layers, suggesting that the application of external pressure should further increase  $T_c$  by enhancing charge transfer between the insulating and conducting layers. The effects of pressure on these iron oxypnictide superconductors may be more prominent than those in the copper-based systems, because the As ion has a greater electronic polarizability, owing to the covalency of the Fe–As chemical bond, and, thus, is more compressible than the divalent  $\text{O}^{2-}$  ion. Here we report that increasing the pressure causes a steep increase in the onset  $T_c$  of F-doped  $\text{LaOFeAs}$ , to a maximum of  $\sim 43$  K at  $\sim 4$  GPa. With the exception of the copper-based high- $T_c$  superconductors, this is the highest  $T_c$  reported to date. The present result, together with the great freedom available in selecting the constituents of isocrystalline materials with the general formula  $\text{LnOTMPn}$  (Ln, Y or rare-earth metal; TM, transition metal; Pn, group-V, ‘pnictogen’, element), indicates that the layered iron oxypnictides are promising as a new material platform for further exploration of high-temperature superconductivity.

$\text{LaOFeAs}$ , a member of the group of quaternary oxypnictides  $\text{LnOTMPn}$ , has a layered structure belonging to the tetragonal  $P4/nmm$  space group with lattice constants  $a = 0.403552(8)$  nm and  $c = 0.87393(2)$  nm. The crystal is composed of a stack of alternating LaO and FeAs layers. It is thought that these two layers are, respectively, positively and negatively charged, and that the La–O chemical bond in the LaO layer is ionic whereas the Fe–As has a predominantly covalent nature. Thus, the chemical formula may be expressed as  $(\text{La}^{3+}\text{O}^{2-})^+(\text{FeAs})^-$ . The conductive carriers, the concentration of which can be increased by substitution of the  $\text{O}^{2-}$  ion by a  $\text{F}^-$  ion, are confined two dimensionally in the  $(\text{FeAs})^-$  layers (Fig. 1). In copper-based high- $T_c$  superconductors, an increase in  $T_c$  has been realized by generation of carriers in a conductive  $\text{CuO}_2$  layer through aliovalent ion doping in the block layer. Similarly, the  $T_c$  of  $\text{LaOFeAs}$  is markedly enhanced on doping of the LaO layer with  $\text{F}^-$  ions. We used two samples of  $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ , one with  $x = 0.05$  (‘under-doped’) and one with  $x = 0.11$  (‘optimally doped’), in this study.

Figure 2a shows the resistivity  $\rho$  of  $\text{LaO}_{0.89}\text{F}_{0.11}\text{FeAs}$  as a function of the temperature  $T$  for pressure  $P < 3$  GPa, determined using a piston–cylinder device. The onset  $T_c$  is determined from the intersection of the two extrapolated lines; one is drawn through the resistivity curve in the normal state just above  $T_c$ , and the other is drawn through the steepest part of the resistivity curve in the superconducting state. This is shown in Fig. 2a for the 0-GPa data. The midpoint  $T_c$  is determined at the temperature where the resistivity is 50% of its value at the onset  $T_c$ . These  $T_c$ s and the zero-point  $T_c$ , determined at zero resistivity, are shown in the inset to Fig. 2a. The onset  $T_c$  increases rapidly with pressure, whereas the zero-point  $T_c$  increases more slowly, resulting in a broadening of the superconducting transition with increasing pressure. This is unlikely to be due to a distribution of pressure in the pressure cell, because the hydrostatic conditions were kept unchanged during the measurements, and is more probably caused by sample inhomogeneity. The sample is a polycrystalline bar cut from a sintered pellet, so there could easily be strains on individual grains that would lead to such a broadening as pressure is increased.

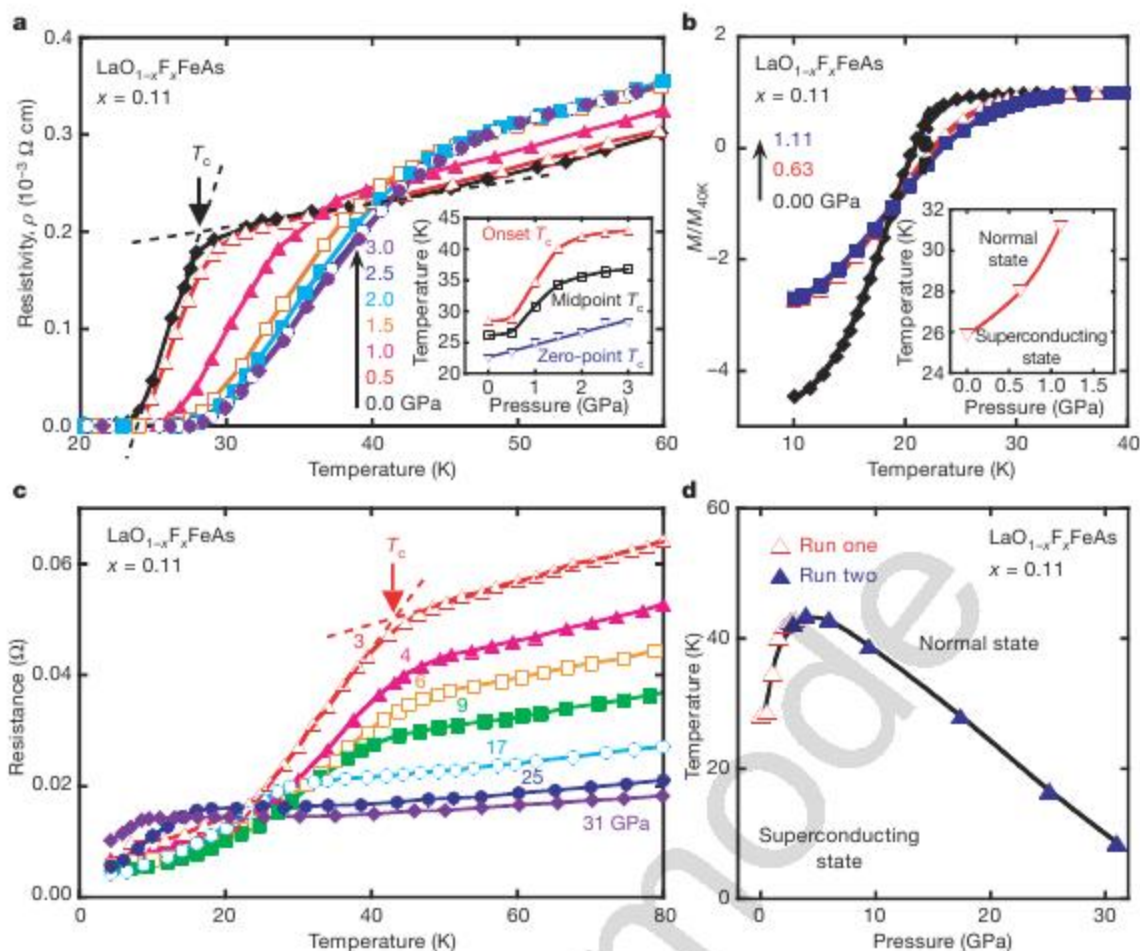
The increase in  $T_c$  with pressure is also confirmed in plots of  $M/M_{40\text{K}}$  (magnetization normalized to its value at 40 K) versus temperature for  $P < 1.1$  GPa (Fig. 2b): here the onset  $T_c$  increases at the same rate as the onset  $T_c$  in the  $\rho$ – $T$  curve (Fig. 2a). However,



**Figure 1 | Schematic crystal structure of  $\text{LaOFeAs}$ .** Electron carriers generated by F-doping into oxygen sites are injected into FeAs metallic layers as a result of the large energy offset between these two layers. We note that the carrier doping layer is spatially separated from the conduction layer.

<sup>1</sup>Department of Physics, College of Humanities and Sciences, Nihon University, Sakurajosui, Setagaya-ku, Tokyo 156-8550, Japan. <sup>2</sup>ERATO-SORST, JST, Frontier Research Center, <sup>3</sup>Frontier Research Center, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama 226-8503, Japan.





**Figure 2 | Superconducting properties in  $\text{LaO}_{0.89}\text{F}_{0.11}\text{FeAs}$  under high pressure.** **a**, Temperature dependence of the electrical resistivity below 3 GPa, using the piston–cylinder device. The onset  $T_c$  and midpoint  $T_c$  values are determined as described in the text. **b**, Temperature dependence of the normalized magnetization  $M/M_{40\text{K}}$  below 1.1 GPa, using a CuBe piston–cylinder device. The magnetization data were obtained by measuring both the sample and the pressure cell. In these measurements, the magnetization of the pressure cell was not subtracted. The pressure dependence of  $T_c$  is shown in the inset, in which the rate of increase of  $T_c$  is similar to that in the resistivity measurements. **c**, Temperature dependence

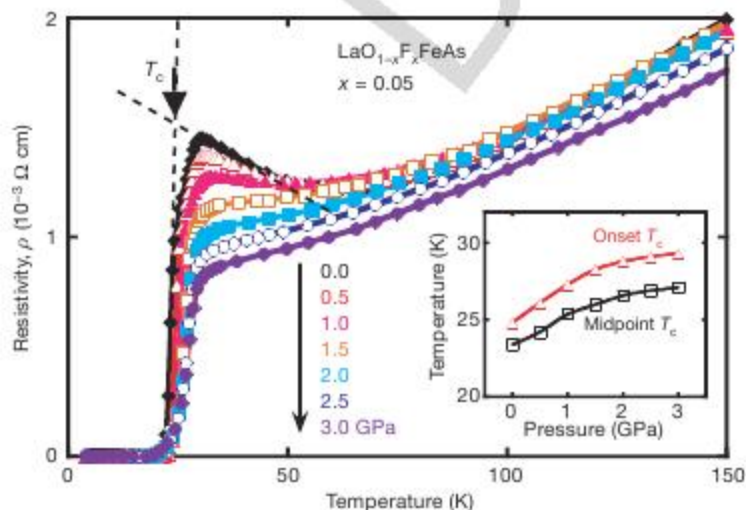
of the electrical resistance above 3 GPa, using the diamond anvil cell. The onset  $T_c$  is determined in the same manner as in **a** (see the text). We find the maximum  $T_c$  of 43 K at 4 GPa, with a decrease to  $T_c = 9$  K at 30 GPa. The superconducting transition is clearly observed in each measurement, although zero resistance is not observed in our measurements down to 4 K. **d**, Superconducting phase diagram of  $\text{LaO}_{0.89}\text{F}_{0.11}\text{FeAs}$ .  $T_c$  is defined at the onset of the transition. Run one (open triangles) shows the electrical resistivity measurements with hydrostatic condition. Run two (filled triangles) is a run using the diamond anvil cell with NaCl as the pressure-transmitting medium. The solid curve is a guide for the eye.

the onset  $T_c$  decreases sharply with further increase in pressure above 3 GPa, as measured using a diamond anvil cell, although zero resistance was not clearly observed in the high-pressure measurements (Fig. 2c). The failure to observe zero resistance is caused by there

being a distribution in the non-hydrostatic compressive stress as a result of the use of a solid pressure-transmitting medium. Figure 2d summarizes the onset  $T_c$  values in the  $P$ – $T$  phase diagram of  $\text{LaO}_{0.89}\text{F}_{0.11}\text{FeAs}$ . Note that  $T_c$  increases steeply as the pressure increases from 1 to 2 GPa, where  $d(\ln T_c)/dP$  is more than 20% per GPa, and then decreases sharply with a further increase in pressure above  $\sim 4$  GPa: the maximum  $T_c$  of 43 K is observed at  $\sim 4$  GPa. The initial increase is much steeper than that observed in high- $T_c$  superconductors, for which the  $d(\ln T_c)/dP$  values are less than 10% per GPa. It is noteworthy that the small resistance loss caused by superconductivity was observed at around 50 K in 4-GPa data (Fig. 2c).

Figure 3 shows the  $\rho$ – $T$  curves for several pressures for the lightly doped compound  $\text{LaO}_{0.95}\text{F}_{0.05}\text{FeAs}$ , which we consider to be an under-doped superconductor, judging from the existence of a resistivity minimum (at  $T_{\min} \approx 60$  K) considerably greater than  $T_c$  in the  $\rho$ – $T$  curves. The value of  $T_{\min}$  decreases continuously with increasing pressure and is zero above  $P = 1$  GPa, whereas  $T_c$  increases gradually with increasing pressure, at a rate of 1.0–2.0 K per GPa (or 3–7% per GPa) in the pressure range 0–3 GPa (Fig. 3 inset). We note that the pressure dependence of  $T_c$  in the under-doped sample is quite different from that in the optimally doped compound  $\text{LaO}_{0.89}\text{F}_{0.11}\text{FeAs}$ .

Various types of pressure effect on the  $T_c$ s of copper-based high- $T_c$  superconductors have been reported<sup>10</sup>. X-ray and neutron diffraction measurements in copper-based superconductors under high pressure have shown that the crystallographic direction normal to the  $\text{CuO}_2$  layer is more compressible than the direction parallel to the



**Figure 3 | Temperature dependence of the electrical resistivity of  $\text{LaO}_{0.95}\text{F}_{0.05}\text{FeAs}$  below 3 GPa, using the piston–cylinder device.** The inset shows the pressure dependence of the onset and midpoint  $T_c$ s. The onset  $T_c$  increases with increasing pressure, with an initial slope of 2.0 K per GPa.



CuO<sub>2</sub> layer; this induces additional changes in the charge distribution, leading to a modulation of the electronic state in the CuO<sub>2</sub> layer. This is one of the reasons why the  $T_c$ s of the copper-based superconductors are further affected by the application of high pressures. In particular, in the case of the spin-ladder compound Ca<sub>14-x</sub>Sr<sub>x</sub>Cu<sub>24</sub>O<sub>41</sub> (ref. 11), the application of pressure induces a metal-insulator transition and supplies sufficient hole carriers to the ladder plane for superconductivity to occur. Doping with F<sup>-</sup> ions, which enhances  $T_c$  in LaOFeAs, causes simultaneous shrinkage of both the *c* and *a* axes. Unlike such 'chemical pressure', an external pressure may induce an anisotropic shrinkage because of the layered structure; this may be the main reason for the further increase in the onset  $T_c$  value up to ~43 K. This also suggests that superconductors with a high  $T_c$  at ambient pressure could be realized among members of the LnOTMPn family by optimizing the chemical composition.

The LnOTMPn family and copper-based superconductors have a common feature in that they both show a magnetic instability in which 3*d*-electron-based carriers with a two-dimensional nature have a critical role. However, there are many differences between these families of superconductors, such as greater covalency of the Fe-As chemical bond in comparison with the Cu-O bond, the stronger two-dimensionality in the iron oxypnictide than in the copper oxide, and there being an odd spin number for Cu<sup>2+</sup> but an even spin number for covalently bonded Fe or Ni. Because LaOCoAs is ferromagnetic with a Curie temperature of 65 K (ref. 12), we suggest that the superconducting phase exists near a magnetic instability, as in the case of Sr<sub>2</sub>RuO<sub>4</sub> (refs 13, 14), hexagonal-close-packed iron<sup>15</sup>, UGe<sub>2</sub> (ref. 16) and so on. These facts imply that the LnOTMPn family can be regarded as a third category of high- $T_c$  superconductor materials if phonon-mediated metal superconductors and copper-based superconductors are defined as the first and second categories, respectively.

In summary, we find iron-based F-doped LaOFeAs to exhibit an onset  $T_c$  of ~43 K under a static pressure of 4 GPa. Other than copper oxide high- $T_c$  superconductors, this is the first 40-K-class superconductor to be identified. Until now, magnetic-element-based materials with a layered structure have been ruled out as a platform for the exploration of high- $T_c$  superconductor materials, but the findings reported here clearly open new avenues of investigation. There are many possible derivatives of the present material system, namely the transition-metal-based, layered oxypnictides, and we expect to make further explorations of these new high- $T_c$  superconductors by varying their chemical composition and studying their behaviour under pressure.

## METHODS SUMMARY

We used dehydrated La<sub>2</sub>O<sub>3</sub> and a mixture powder of LaAs, Fe<sub>2</sub>As and FeAs as starting materials, and synthesized the target materials using the following solid-state reactions in a sealed silica tube.

To obtain the mixed powder of LaAs, Fe<sub>2</sub>As and FeAs, we mixed metallic La, metallic Fe and metallic As in a ratio of 1:3:3 in an Ar gas, sealed it in a silica tube under pumping, and subsequently heated the mixture at 850 °C for ten hours. Because the vapour pressure of elementary As is high at temperatures >600 °C, the synthesis of the As compounds as the precursor was required to avoid the explosion of the sealed glass sample during the main reactions, as follows. The La<sub>2</sub>O<sub>3</sub> that we used was obtained by heating commercially available La<sub>2</sub>O<sub>3</sub> powders at 600 °C for ten hours in air to remove adsorbed water. Then a 1:1 mixture of the LaAs-Fe<sub>2</sub>As-FeAs powder and the La<sub>2</sub>O<sub>3</sub> powder was compressed at 10 MPa into pellets either of 7-mm diameter and 5-mm thickness or of dimensions 7 mm by 10 mm by 10 mm. A pellet was placed in a silica tube, and sealed with 0.02-MPa Ar gas. The sealed silica tube was heated at 1,250 °C for 40 hours

to synthesize undoped LaOFeAs. The Ar gas was used to prevent collapse of the silica tube during this heating process. For F-doping, some of the La<sub>2</sub>O<sub>3</sub> was replaced with a 1:1 mixture of LaF<sub>3</sub> and metallic La in the starting materials for undoped LaOFeAs.

Electrical resistivity measurements under high pressure were performed by means of a standard direct-current four-probe method. Pressures of up to 3 GPa were applied and clamped at room temperature using a piston (WC)-cylinder (NiCrAl) device<sup>17</sup>. The dimensions of the samples, cut from a sintered pellet, were 2.5 × 1.5 × 1.0 mm<sup>3</sup> and 3.3 × 1.3 × 0.5 mm<sup>3</sup> for *x* = 0.11 and 0.05, respectively. A liquid pressure-transmitting medium (Fluorinert liquids FC-70 and FC-77 in a 1:1 ratio) was used to maintain hydrostatic conditions. A diamond anvil cell made of CuBe alloy was used for electrical resistance measurements at pressures up to 30 GPa: in this case, the sample chamber (a rhenium gasket) was filled with powdered NaCl for the pressure-transmitting medium, and thin (10-μm-thick) platinum ribbons were inserted into the sample chamber to act as leads for the standard direct-current four-probe analysis. The dimensions of the rectangular sample were 0.1 × 0.1 × 0.03 mm<sup>3</sup>. A thin BN layer acted as an electric insulator between the leads and the rhenium gasket. Fine ruby powder scattered in the sample chamber was used to determine the pressure by means of a standard ruby fluorescence method. The high-pressure magnetization measurements were performed using a superconducting-quantum-interference-device magnetometer (Quantum Design MPMS) used in conjunction with a piston-cylinder device made of CuBe and a liquid pressure-transmitting medium (FC-70:FC-77 = 1:1).

Received 26 February; accepted 8 April 2008.

Published online 23 April 2008.

- Kamihara, Y. et al. Iron-based layered superconductor: LaOFeP. *J. Am. Chem. Soc.* **128**, 10012–10013 (2006).
- Liang, C. Y. et al. Synthesis and structural characterization of LaOFeP superconductors. *Supercond. Sci. Technol.* **20**, 687–690 (2007).
- Watanabe, T. et al. Nickel-based oxyphosphide superconductor with a layered crystal structure, LaNiOP. *Inorg. Chem.* **46**, 7719–7721 (2007).
- Kamihara, Y., Watanabe, T., Hirano, M. & Hosono, H. Iron-based layered superconductor La[O<sub>1-x</sub>F<sub>x</sub>]FeAs (*x* = 0.05–0.12) with  $T_c$  = 26 K. *J. Am. Chem. Soc.* **130**, 3296–3297 (2008).
- Kohama, Y. et al. Ferromagnetic spin fluctuation in LaFeAsO<sub>1-x</sub>F<sub>x</sub>. *Phys. Rev. Lett.* (submitted).
- Nakai, Y., Ishida, K., Kamihara, Y., Hirano, M. & Hosono, H. Spin dynamics in iron-based layered superconductor La<sub>0.87</sub>Ca<sub>0.13</sub>OFeP revealed by <sup>31</sup>P and <sup>139</sup>La NMR studies. *Phys. Rev. Lett.* (submitted).
- Ishibashi, S., Terakura, K. & Hosono, H. A possible ground state and its electronic structure of a mother material (LaOFeAs) of new superconductors. *J. Phys. Soc. Jpn* (in the press).
- Kuroki, K. et al. Unconventional superconductivity originating from disconnected Fermi surfaces in LaO<sub>1-x</sub>F<sub>x</sub>FeAs. Preprint at (<http://arxiv.org/abs/0803.3325>) (2008).
- Singh, D. J. & Du, M.-H. LaFeAsO<sub>1-x</sub>F<sub>x</sub>: A low carrier density superconductor near itinerant magnetism. Preprint at (<http://arxiv.org/abs/0803.0429>) (2008).
- Takahashi, H. & Mori, N. in *Studies of High Temperature Superconductors* Vol. 16 (ed. Narlikar, A.) 1–63 (Nova Science, New York, 1996).
- Uehara, M. et al. Superconductivity in the spin ladder system Sr<sub>0.4</sub>Ca<sub>13.6</sub>Cu<sub>24</sub>O<sub>41.84</sub>. *J. Phys. Soc. Jpn.* **65**, 2764–2767 (1996).
- Yanagi, H. et al. Itinerant ferromagnetism in layered crystals LaCoOPn (*Pn* = P, As). *Phys. Rev. B* (submitted).
- Maeno, Y. et al. Superconductivity in a layered perovskite without copper. *Nature* **372**, 532–534 (1994).
- Ishida, K. et al. Spin-triplet superconductivity in Sr<sub>2</sub>RuO<sub>4</sub> identified by <sup>17</sup>O Knight shift. *Nature* **396**, 658–660 (1998).
- Shimizu, K. et al. Superconductivity in the non-magnetic state of iron under pressure. *Nature* **412**, 316–318 (2001).
- Saxena, S. S. et al. Superconductivity on the border of itinerant-electron ferromagnetism in UGe<sub>2</sub>. *Nature* **406**, 587–592 (2000).
- Matsumoto, T. Materials for high-pressure apparatus. *Rev. High Pressure Sci. Technol.* **12**, 280–287 (2002).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.T. ([hiroki@chs.nihon-u.ac.jp](mailto:hiroki@chs.nihon-u.ac.jp)).



# High-resolution carbon dioxide concentration record 650,000–800,000 years before present

Dieter Lüthi<sup>1</sup>, Martine Le Floch<sup>2</sup>, Bernhard Bereiter<sup>1</sup>, Thomas Blunier<sup>1</sup>†, Jean-Marc Barnola<sup>2</sup>, Urs Siegenthaler<sup>1</sup>, Dominique Raynaud<sup>2</sup>, Jean Jouzel<sup>3</sup>, Hubertus Fischer<sup>4</sup>, Kenji Kawamura<sup>1</sup>† & Thomas F. Stocker<sup>1</sup>

Changes in past atmospheric carbon dioxide concentrations can be determined by measuring the composition of air trapped in ice cores from Antarctica. So far, the Antarctic Vostok and EPICA Dome C ice cores have provided a composite record of atmospheric carbon dioxide levels over the past 650,000 years<sup>1–4</sup>. Here we present results of the lowest 200 m of the Dome C ice core, extending the record of atmospheric carbon dioxide concentration by two complete glacial cycles to 800,000 yr before present. From previously published data<sup>1–8</sup> and the present work, we find that atmospheric carbon dioxide is strongly correlated with Antarctic temperature throughout eight glacial cycles but with significantly lower concentrations between 650,000 and 750,000 yr before present. Carbon dioxide levels are below 180 parts per million by volume (p.p.m.v.) for a period of 3,000 yr during Marine Isotope Stage 16, possibly reflecting more pronounced oceanic carbon storage. We report the lowest carbon dioxide concentration measured in an ice core, which extends the pre-industrial range of carbon dioxide concentrations during the late Quaternary by about 10 p.p.m.v. to 172–300 p.p.m.v.

Two deep ice cores have been drilled within the European Project for Ice Coring in Antarctica (EPICA). The drillings at both sites, one located in the Dronning Maud Land (DML) area at Kohnen Station (75° 00' 06" S; 00° 04' 04" E), the other at Concordia Station (Dome C; 75° 06' 04" S; 123° 20' 52" E), have been stopped at or a few metres above bedrock at a depth of 2,774 m and 3,270 m, respectively. EPICA produced climate records focusing, among others, on water isotopes, particulate and soluble aerosol species as well as greenhouse gases<sup>4,8,9</sup>. For the latter, the analysis of the air entrapped in the ice is the only direct way to determine their concentrations for times before precise routine atmospheric measurements were done, that is, before 1958.

Carbon dioxide measurements on Dome C ice, focusing on the interval 390 to 650 kyr before present, BP (2,700–3,060 m)<sup>4</sup>, confirmed the strong coupling between CO<sub>2</sub> and Antarctic temperature found<sup>1</sup> in the Vostok ice core for the past 420 kyr. They also showed a range in CO<sub>2</sub> about 30% smaller (180 to 260 p.p.m.v.) than for the Vostok interval (180 to 300 p.p.m.v.) related to cooler interglacials in the earlier period<sup>10</sup>.

Investigating the lowest 200 m of the Dome C core (3,060–3,270 m), we extend the existing CO<sub>2</sub> records by two complete glacial cycles. The data set of the depth interval 3,060–3,190 m, to which two laboratories have contributed with three different air extraction techniques and two independent measurement methods, is shown in Fig. 1. The resolution of the samples extracted by a needle cracker

and analysed by laser absorption spectroscopy (LAS) at the University of Bern averages 0.55 m resulting in a mean time resolution of 570 yr. The 47 values measured at LGGE in Grenoble (ball crusher and gas chromatography) covering depth levels around 3,062 m, between 3,085 and 3,095 m, and below 3,120 m, average slightly higher than neighbouring samples analysed at Bern (see Supplementary Information). Despite this difference, which seems to be larger for low concentrations, the Grenoble measurements confirm the shape of the Bern record. Further confidence in the accuracy of the record is given by LAS measurements of air extracted by sublimating the ice at selected depth levels (Fig. 1). Also, remeasurements of depths between 3,030 m and 3,060 m (615–665 kyr BP) agree with earlier data<sup>4</sup> within the experimental uncertainty (Fig. 2).

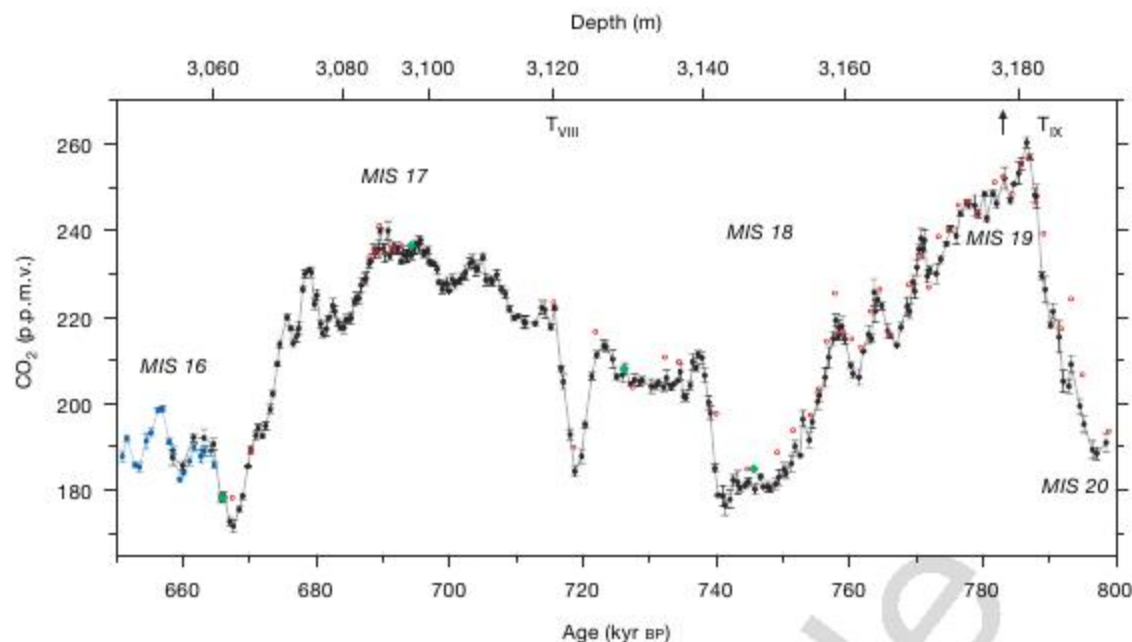
At a depth level of 3,178 m, a CO<sub>2</sub> value of 339 p.p.m.v. (average of four samples) with a large scattering between 264 and 477 p.p.m.v. is found (black arrow, Fig. 1). This artefact can be explained by the fact that this ice is from an ice-core section drilled towards the end of the season 2002–2003, when an ethanol–water mixture had to be added at the bottom of the borehole to allow further drilling. This caused partial melting at the outside of the core and subsequent refreezing when hoisting the core through colder sections of the bore hole. This is the only artefact discovered in the ice above 3,190.56 m.

The part below 3,190 m is characterized by CO<sub>2</sub> concentrations mainly fluctuating between 180 and 210 p.p.m.v. in both the LGGE and the Bern data. Because of disturbed ice stratigraphy<sup>8</sup>, this part of the core does not reflect a climatic signal. We therefore limit further discussion of our new data set to the shallower part above 3,190.56 m, that is, back to the glacial corresponding to Marine Isotope Stage (MIS) 20, 800 kyr BP.

The main features of the record are as follows. Starting at a glacial level of 188 p.p.m.v. in MIS 20, the CO<sub>2</sub> concentration rises by about 70 p.p.m.v. within 9 kyr. This slope is similar to the five glacial–interglacial transitions of the past 450 kyr (see Supplementary Information for a detailed comparison). At the beginning of MIS 19, CO<sub>2</sub> attains its local maximum of 261 p.p.m.v., which is similar to the highest levels during MIS 13 and 15 but substantially lower than during the four recent interglacials. Similar to MIS 5 and 9, CO<sub>2</sub> drops by about 15 p.p.m.v. in MIS 19 to a slowly decreasing phase of about 10 kyr, before it declines in a sawtooth pattern to glacial levels. This slow decline in the 10 kyr after the initial peak is in contrast to the slowly increasing CO<sub>2</sub> levels in the Holocene and the even slower increase in MIS 11, despite the similar orbital configuration of these three periods.

<sup>1</sup>Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland, and Oeschger Centre for Climate Change Research, University of Bern, Erlachstrasse 9a, CH-3012 Bern, Switzerland. <sup>2</sup>Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), CNRS-Université Joseph Fourier Grenoble, 54 Rue Molière, 38402 St Martin d'Hères, France. <sup>3</sup>Institut Pierre Simon Laplace/Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-Université Versailles-Saint Quentin, CE Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France. <sup>4</sup>Alfred Wegener Institute for Polar and Marine Research (AWI), Columbusstrasse, D-27568 Bremerhaven, Germany. †Present addresses: Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen OE, Denmark (T.B.); National Institute of Polar Research, Research Organization of Information and Systems, 1-9-10 Kaga, Itabashi-ku, Tokyo 173-8515, Japan (K.K.).





**Figure 1 | Dome C CO<sub>2</sub> data.** Black solid circles connected by a grey line: data measured at Bern (mean of four to six samples); error bars represent  $1\sigma$  of the mean (s.d.). Red open circles: data measured at Grenoble ( $2\sigma = 3$  p.p.m.v.). Blue solid circles: Dome C CO<sub>2</sub> data published by Siegenthaler *et al.*<sup>4</sup>. Green solid diamonds: control measurements with the

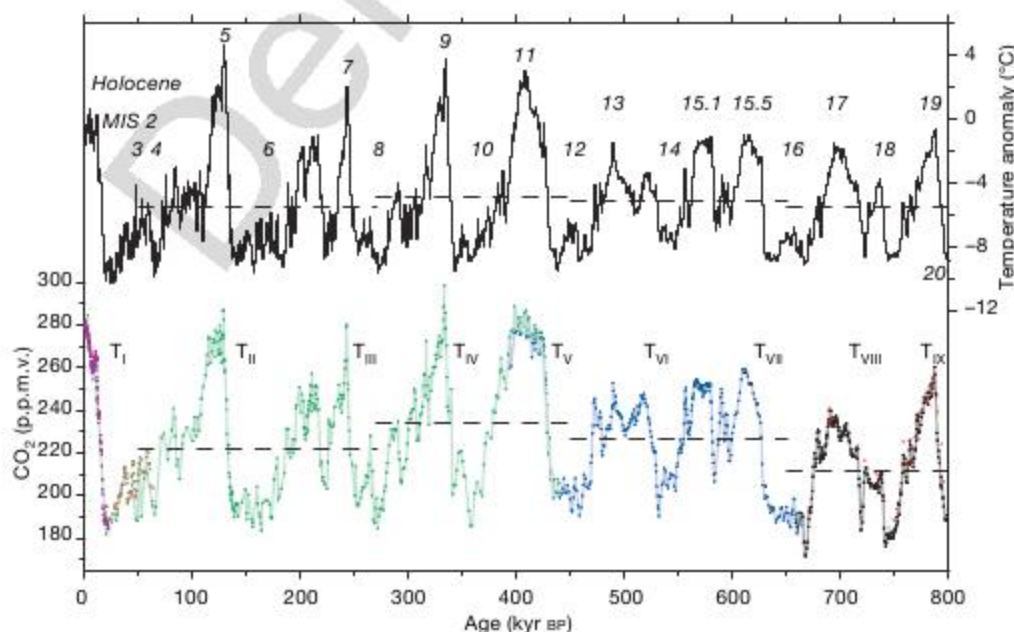
sublimation extraction technique. The black arrow indicates a CO<sub>2</sub> value of  $339 \pm 56$  p.p.m.v. (s.d.), an artefact detected at a depth of 3,178.12 m (age: 783,040 yr BP). All data are plotted on the EDC3\_gas\_a age scale<sup>26</sup>. Glacial terminations are indicated using Roman numerals in subscript (for example T<sub>VIII</sub>); MIS denotes Marine Isotope Stage<sup>27</sup>.

MIS 18 consists of two clearly separated phases. The earlier phase reaches its minimum of 177 p.p.m.v. just before a fast increase to the second, a phase of rather constant CO<sub>2</sub> concentration between 205 and 212 p.p.m.v. lasting for 20 kyr. A rapid reduction by about 30 p.p.m.v. within a few thousand years, similar to MIS 15.2 and 15.4, and a similarly rapid increase by 40 p.p.m.v. (termination VIII) mark the beginning of the next interglacial. During the 40,000 yr of MIS 17, CO<sub>2</sub> ranges between 215 and 240 p.p.m.v., which is significantly lower than in other interglacials during the past 800 kyr. At the beginning of MIS 16, CO<sub>2</sub> remains below 180 p.p.m.v. for 3 kyr, most probably reflecting more pronounced glacial carbon storage in the ocean. During this period, CO<sub>2</sub> falls to its lowest value ever found in ice cores, 172 p.p.m.v. (667 kyr BP),

redefining the natural range of CO<sub>2</sub> of the late Quaternary to about 170 to 300 p.p.m.v., before it rises at a rate of 8 p.p.m.v. kyr<sup>-1</sup> to 190 p.p.m.v. at 665 kyr BP.

Figure 2 shows our data together with earlier results from the Dome C (650–390 kyr BP<sup>4</sup> and 22–0 kyr BP<sup>5</sup>), Vostok<sup>1–3</sup> (440–0 kyr BP) and Taylor Dome<sup>6</sup> (60–20 kyr BP) ice cores resulting in a composite CO<sub>2</sub> record over eight glacial cycles. During these 800 kyr, CO<sub>2</sub> is strongly coupled with the Antarctic temperature ( $r^2 = 0.82$ ).

It was suggested earlier<sup>4</sup> that there is a strong stationary relationship between Antarctic temperature and CO<sub>2</sub>. But our data reveal a significant deviation from this behaviour: The atmospheric concentration of CO<sub>2</sub> during MIS 17 remains significantly below the levels



**Figure 2 | Compilation of CO<sub>2</sub> records and EPICA Dome C temperature anomaly over the past 800 kyr.** The Dome C temperature anomaly record with respect to the mean temperature of the last millennium<sup>8</sup> (based on original deuterium data interpolated to a 500-yr resolution), plotted on the EDC3 timescale<sup>13</sup>, is given as a black step curve. Data for CO<sub>2</sub> are from Dome C (solid circles in purple<sup>5</sup>, blue<sup>4</sup>, black: this work, measured at Bern; red open

circles: this work, measured at Grenoble), Taylor Dome<sup>6</sup> (brown) and Vostok<sup>1–3</sup> (green). All CO<sub>2</sub> values are on the EDC3\_gas\_a age scale<sup>26</sup>. Horizontal lines are the mean values of temperature and CO<sub>2</sub> for the time periods 799–650, 650–450, 450–270 and 270–50 kyr BP. Glacial terminations are indicated using Roman numerals in subscript (for example T<sub>I</sub>); Marine Isotope Stages (MIS) are given in italic Arabic numerals<sup>27</sup>.

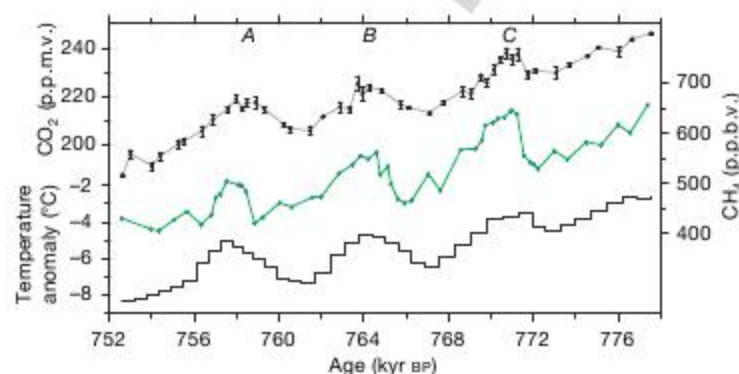


during MIS 13, 15 and 19; this is expected neither from the temperature variations which always reach comparable levels during these interglacials nor from carbon cycle models<sup>11</sup>. Except for MIS 14, the temperature anomalies (relative to the mean temperature of the last millennium) of the coldest levels of all glacial periods range from around  $-9$  to  $-9.5$  °C with CO<sub>2</sub> concentrations generally in the range 180–190 p.p.m.v.. Exceptions are MIS 16 and MIS 18 where we find significantly lower concentrations of 172–180 p.p.m.v.. To illustrate this difference in the CO<sub>2</sub>/temperature relationship before and after MIS 16, we calculate mean values for four different time intervals (799–650, 650–450, 450–270 and 270–50 kyr BP). All intervals start and end in the middle of a glacial including two interglacials (see Supplementary Information for details). Whereas averaged temperature anomalies of these four intervals ( $-5.5$ ,  $-5.1$ ,  $-4.9$  and  $-5.5$  °C, respectively) only slightly differ from one another, the mean value of CO<sub>2</sub> between 799 and 650 kyr (212 p.p.m.v.) is 4–9% lower than for the ensuing periods (227, 234 and 222 p.p.m.v.; see further illustrations in the Supplementary Information).

This shift is unexpected and takes place where the new measurement interval for this study begins. However, it seems to be a robust feature and not a measurement artefact, as we find no alteration in our CO<sub>2</sub> record due to (1) enlarged amounts of impurities<sup>12</sup>, (2) increasing temperature with depth ( $-7.2$  °C at 3060 m to  $-4.2$  °C at 3,200 m, pressure melting point:  $-2.3$  °C) or (3) changes in the ice structure. Also, we observe no peculiarities in the preliminary O<sub>2</sub>/N<sub>2</sub> and air content data (D. Raynaud and G. Dreyfus, personal communication) during this period. Furthermore, there is an overall good agreement between the data from Siegenthaler *et al.*<sup>4</sup>, the Grenoble and the Bern measurements (this work) using different extraction techniques. Taken together, we conclude that our record reflects the true atmospheric CO<sub>2</sub> concentration.

Instead of CO<sub>2</sub> concentrations being about 15 p.p.m.v. lower, the CO<sub>2</sub>/temperature shift can also be explained by isotopic composition overestimating temperature by about 1 °C. However, both the elevation<sup>13</sup> and moisture origin corrections on the estimated temperature change do not indicate a systematic overestimation of temperature between MIS 16 and 18.

This implies that there could be a long-term CO<sub>2</sub> increase by about 25 p.p.m.v. from 800 to 400 kyr BP. Together with the long-term decrease of 15 p.p.m.v. during the past four glacial cycles, we suggest significant slow fluctuations in the atmospheric CO<sub>2</sub> concentration on timescales of several 10<sup>5</sup> years, probably influenced by changes in the weathering<sup>14</sup> or by major reorganizations in the carbon reservoir of the global ocean<sup>15</sup>. Older ice is required to test hypotheses about any consistent long-term CO<sub>2</sub> trend and a probable connection with the 500-kyr variations of the Quaternary carbon cycle found in marine records<sup>15</sup>.



**Figure 3 | AIM events during MIS 18.** Variations of Dome C CO<sub>2</sub> (black solid circles; mean of four to six samples; error bars are 1 $\sigma$  of the mean (s.d.)), methane<sup>24</sup> (green) and temperature anomaly based on the mean temperature of the past millennium<sup>8</sup> (black step curve) between 778 and 752 kyr BP. Gases are plotted on the EDC3\_gas age scale<sup>26</sup>, temperature on the EDC3 timescale<sup>13</sup>. The letters A to C denote the three events.

Besides long-term changes, the response of the carbon cycle to millennial climate change is of special interest. Indermühle *et al.*<sup>6</sup> found four well-defined oscillations with a peak-to-peak amplitude of about 15 p.p.m.v. during the last glacial in the Taylor Dome ice core. These variations are concomitant to the larger of the Antarctic Isotope Maximum (AIM) events (Antarctic warming events with an amplitude of 1–3 °C)<sup>7,16,17</sup>. AIM events are counterparts of the short and more pronounced warming events in the Northern Hemisphere known as Dansgaard–Oeschger events<sup>18</sup>, and hence are a manifestation of the bipolar seesaw<sup>19–21</sup> with oceanic and terrestrial processes likely contributing to the reconstructed CO<sub>2</sub> variations<sup>22,23</sup>.

The oldest trace of millennial-scale variations in our record is detected between 778 and 752 kyr BP (Fig. 2). Antarctic temperature<sup>8</sup>, CO<sub>2</sub> and the atmospheric concentration of methane<sup>24</sup>, whose variations are primarily affected by changes in Northern Hemisphere temperature and precipitation<sup>25</sup>, show three well-defined relative maxima (Fig. 3). A comparison of these data with the corresponding variations during MIS 3 reveals that all three quantities vary in the same range during both periods (see Supplementary Table 3). Furthermore, the rates of increase of CH<sub>4</sub> for these three events in MIS 18 are similar to those in the EDC ice core during the large AIM events of MIS 3. Finally, the gradual rise of CO<sub>2</sub> for events A and B precedes the rapid rise of CH<sub>4</sub> by a few millennia, which is another typical feature of large AIM events in MIS 3<sup>7</sup>. Although the phasing for event C is less clear because of the relatively small CO<sub>2</sub> increase, our findings provide strong support for an active bipolar seesaw during MIS 18.

## METHODS SUMMARY

The analytical methods for routine CO<sub>2</sub> measurements used at the University of Bern and at LGGE in Grenoble are based on dry extraction techniques followed by laser absorption spectroscopy and gas chromatography, respectively. In Bern, four to six samples (23 × 23 × 16 mm; 8 g) from each depth level (every 550 mm) in the ice core are measured in random order (two on the same day; the following two after all depth intervals have been measured twice). The samples are crushed by a cooled needle cracker under vacuum conditions. The sample container is connected to a cold trap for several minutes to release the air from the clathrates. Afterwards, the air is expanded to a measuring cell, where a laser is tuned six times over the absorption line of a vibration–rotation transition of the CO<sub>2</sub> molecule. The calibration is done using a CO<sub>2</sub> in air standard gas (251.65 p.p.m.v.) scaled on the World Meteorological Organization mole fraction scale.

In Grenoble, on every depth level one to three samples of about 40 g of ice are crushed under vacuum conditions. About 20 min later, the extracted gas is expanded in the sample loop of the gas chromatograph and analysed. Depending on the amount of the extracted air, three to five successive analyses are done. To avoid the possible influence of the water vapour injected with the gas, the CO<sub>2</sub> ratio is calculated as the ratio between the CO<sub>2</sub> peak and the air (O<sub>2</sub> + N<sub>2</sub>) peak. The calibration is done using an Air Liquide standard scaled on three CSIRO standards (260.3 p.p.m.v., 321.1 p.p.m.v. and 172.8 p.p.m.v.).

Received 12 October 2007; accepted 17 March 2008.

- Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Pepin, L., Raynaud, D., Barnola, J. M. & Loutre, M. F. Hemispheric roles of climate forcings during glacial–interglacial transitions as deduced from the Vostok record and LLN-2D model experiments. *J. Geophys. Res.* **106**, 31885–31892 (2001).
- Raynaud, D. *et al.* The record for marine isotopic stage 11. *Nature* **436**, 39–40 (2005).
- Siegenthaler, U. *et al.* Stable carbon cycle–climate relationship during the Late Pleistocene. *Science* **310**, 1313–1317 (2005).
- Monnin, E. *et al.* Atmospheric CO<sub>2</sub> concentrations over the last glacial termination. *Science* **291**, 112–114 (2001).
- Indermühle, A., Monnin, E., Stauffer, B., Stocker, T. F. & Wahlen, M. Atmospheric CO<sub>2</sub> concentration from 60 to 20 kyr BP from the Taylor Dome ice core, Antarctica. *Geophys. Res. Lett.* **27**, 735–738 (2000).
- Ahn, J. H. & Brook, E. J. Atmospheric CO<sub>2</sub> and climate from 65 to 30 ka BP. *Geophys. Res. Lett.* **34**, L10703, doi:10.1029/2007/GL029551 (2007).
- Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the last 800,000 years. *Science* **317**, 793–796 (2007).
- Wolff, E. W. *et al.* Southern Ocean sea ice extent, productivity and iron flux over the past eight glacial cycles. *Nature* **440**, 491–496 (2006).



10. EPICA Community Members. Eight glacial cycles from an Antarctic ice core. *Nature* **429**, 623–628 (2004).
11. Köhler, P. & Fischer, H. Simulating low frequency changes in atmospheric CO<sub>2</sub> during the last 740 000 years. *Clim. Past* **2**, 57–78 (2006).
12. Lambert, F. *et al.* Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452**, 616–619 (2008).
13. Parrenin, F. *et al.* The EDC3 chronology for the EPICA Dome C ice core. *Clim. Past* **3**, 485–497 (2007).
14. Kump, L. R., Brantley, S. L. & Arthur, M. A. Chemical weathering, atmospheric CO<sub>2</sub>, and climate. *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).
15. Wang, P., Tian, J., Cheng, X., Liu, C. & Xu, J. Major Pleistocene stages in a carbon perspective: The South China Sea record and its global comparison. *Paleoceanography* **19**, PA4005, doi:10.1029/2003PA000991 (2004).
16. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
17. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
18. North Greenland Ice Core Project Members. High-resolution climate record of the Northern Hemisphere back into the last interglacial period. *Nature* **431**, 147–151 (2004).
19. Broecker, W. S. Paleocirculation during the last deglaciation: a bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
20. Stocker, T. F. The seesaw effect. *Science* **282**, 61–62 (1998).
21. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18**, 1087, doi:10.1029/2003PA000920 (2003).
22. Marchal, O. *et al.* Modelling the concentration of atmospheric CO<sub>2</sub> during the Younger Dryas climate event. *Clim. Dyn.* **15**, 341–354 (1999).
23. Köhler, P., Joos, F., Gerber, S. & Knutti, R. Simulated changes in vegetation distribution, land carbon storage, and atmospheric CO<sub>2</sub> in response to a collapse of the North Atlantic thermohaline circulation. *Clim. Dyn.* **25**, 689–708 (2005).
24. Louergue, L. *et al.* Orbital and millennial-scale features of atmospheric CH<sub>4</sub> over the past 800,000 years. *Nature* doi:10.1038/nature06950 (this issue).
25. Siddall, M. *et al.* Using a maximum simplicity paleoclimate model to simulate millennial variability during the last four glacial periods. *Quat. Sci. Rev.* **25**, 3185–3197 (2006).
26. Louergue, L. *et al.* New constraints on the gas age-ice age difference along the EPICA ice cores, 0–50 kyr. *Clim. Past* **3**, 527–540 (2007).
27. Lisiecki, L. E. & Raymo, M. E. A. Pliocene–Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, PA1003 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work is a contribution to the European Project for Ice Coring in Antarctica (EPICA), a joint European Science Foundation/European Commission scientific program, funded by the European Commission and by national contributions from Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom. The main logistic support was provided by IPEV and PNRA (at Dome C). We thank the technical team on the field and at both laboratories, G. Hausammann for helping with the CO<sub>2</sub> measurements, and B. Stauffer, J. Schwander, M. Leuenberger, F. Joos, V. Masson-Delmotte, G. Dreyfus and C. Körner for their input. We acknowledge financial support by the Swiss NSF, the University of Bern, the Swiss Federal Agency of Energy and the French ANR (Agence nationale pour la Recherche; programme PICC). This is EPICA publication no. 194.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.L. (luethi@climate.unibe.ch).

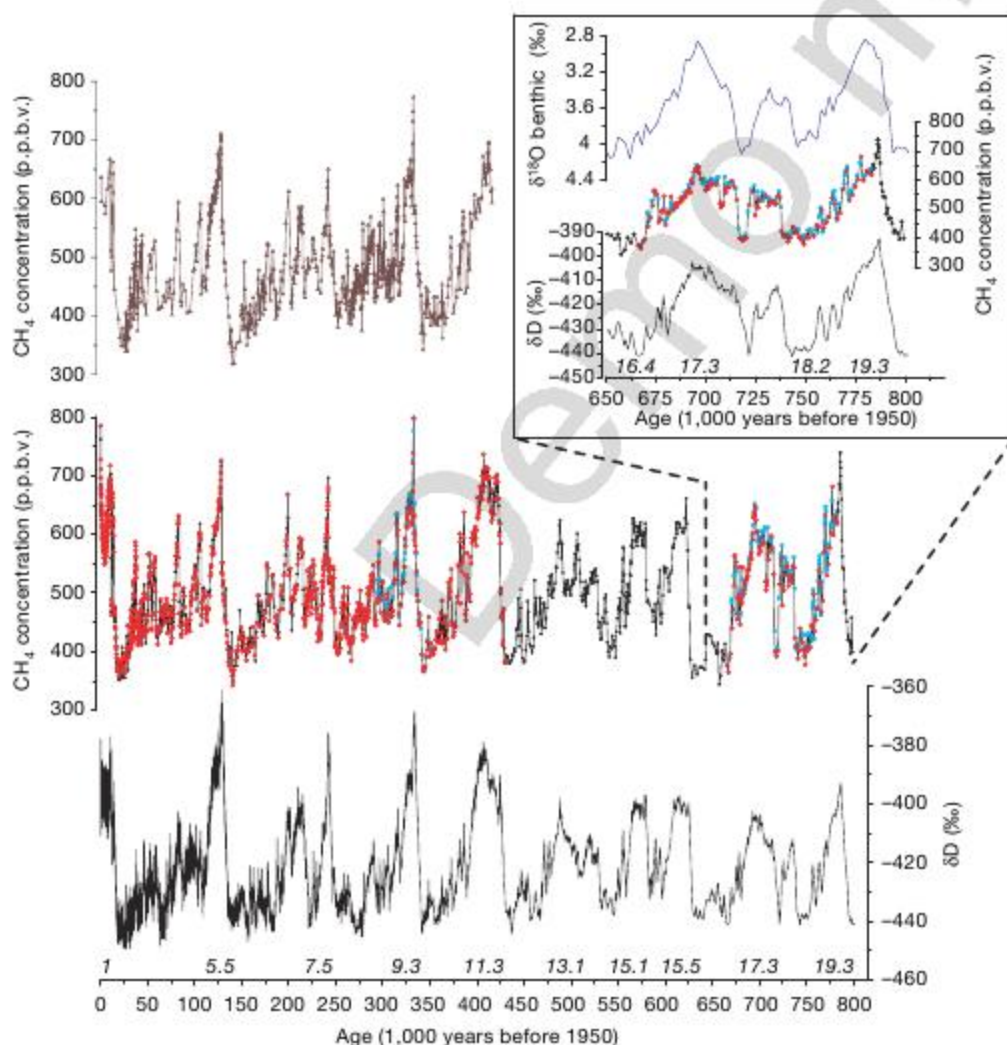


# Orbital and millennial-scale features of atmospheric CH<sub>4</sub> over the past 800,000 years

Laetitia Louergue<sup>1</sup>, Adrian Schilt<sup>2</sup>, Renato Spahni<sup>2†</sup>, Valérie Masson-Delmotte<sup>3</sup>, Thomas Blunier<sup>2‡</sup>, Bénédicte Lemieux<sup>1</sup>, Jean-Marc Barnola<sup>1</sup>, Dominique Raynaud<sup>1</sup>, Thomas F. Stocker<sup>2</sup> & Jérôme Chappellaz<sup>1</sup>

Atmospheric methane is an important greenhouse gas and a sensitive indicator of climate change and millennial-scale temperature variability<sup>1</sup>. Its concentrations over the past 650,000 years have varied between ~350 and ~800 parts per 10<sup>9</sup> by volume (p.p.b.v.) during glacial and interglacial periods, respectively<sup>2</sup>. In comparison, present-day methane levels of ~1,770 p.p.b.v. have been reported<sup>3</sup>. Insights into the external forcing factors and internal feedbacks controlling atmospheric methane are essential for predicting the methane budget in a warmer world<sup>3</sup>. Here we present a detailed atmospheric methane record from the EPICA Dome C ice core that extends the history of this greenhouse gas

to 800,000 yr before present. The average time resolution of the new data is ~380 yr and permits the identification of orbital and millennial-scale features. Spectral analyses indicate that the long-term variability in atmospheric methane levels is dominated by ~100,000 yr glacial–interglacial cycles up to ~400,000 yr ago with an increasing contribution of the precessional component during the four more recent climatic cycles. We suggest that changes in the strength of tropical methane sources and sinks (wetlands, atmospheric oxidation), possibly influenced by changes in monsoon systems and the position of the intertropical convergence zone, controlled the atmospheric methane budget, with an



**Figure 1 | Methane records and EPICA/Dome C  $\delta D$ .** Bottom to top:  $\delta D$  record<sup>9</sup>; EDC methane record (previously published data<sup>2</sup>, black diamonds; new data from LGGE, red diamonds; new data from Bern, blue dots); Vostok methane record<sup>1</sup>. Marine Isotope Stage numbering is given at the bottom of each interglacial. Insert: expanded view of the bottom section of EDC:  $\delta D$  values<sup>9</sup> (black line), CH<sub>4</sub> (black line) from EDC and stack benthic  $\delta^{18}O$  values (blue line)<sup>19</sup> for the period from MIS 16 to 20.2, on their respective age scales.  $\delta^{18}O = [(^{18}O/^{16}O)_{\text{sample}} / (^{18}O/^{16}O)_{\text{standard}}] - 1$ , where standard is vPDB;  $\delta D = [(D/H)_{\text{sample}} / (D/H)_{\text{standard}}] - 1$  where standard is SMOW.

<sup>1</sup>Laboratoire de Glaciologie et Géophysique de l'Environnement, CNRS-Université Joseph Fourier Grenoble, 54 Rue Molière, 38402 St Martin d'Hères, France. <sup>2</sup>Climate and Environmental Physics, Physics Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland, and Oeschger Centre for Climate Change Research, University of Bern, Erlachstrasse 9a, CH-3012 Bern, Switzerland. <sup>3</sup>Institut Pierre Simon Laplace/Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-University Versailles-Saint Quentin, CE Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette, France. <sup>†</sup>Present addresses: Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, United Kingdom (R.S.); Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen OE, Denmark (T.B.).



additional source input during major terminations as the retreat of the northern ice sheet allowed higher methane emissions from extending periglacial wetlands. Millennial-scale changes in methane levels identified in our record as being associated with Antarctic isotope maxima events<sup>1,4</sup> are indicative of ubiquitous millennial-scale temperature variability during the past eight glacial cycles.

Atmospheric methane is an important climate forcing as well as a sensitive indicator of climate change and variability. Over the past 650 kyr, the ice-core record indicates that its abundance has varied from ~350 p.p.b.v. during glacial periods up to ~800 p.p.b.v. during interglacials<sup>2</sup>. In 2005, its global average was  $1,774 \pm 1.8$  p.p.b.v. (ref. 3), far exceeding the natural range of the past 650 kyr. Understanding the link between external forcings and internal feedbacks on the natural CH<sub>4</sub> budget is important for forecasting the latter in a warmer world. In addition, CH<sub>4</sub> is tightly linked to Northern Hemisphere millennial variability during the last glacial period<sup>4</sup>. Therefore, CH<sub>4</sub> variations can be used as a proxy for abrupt change further back in time<sup>1</sup>.

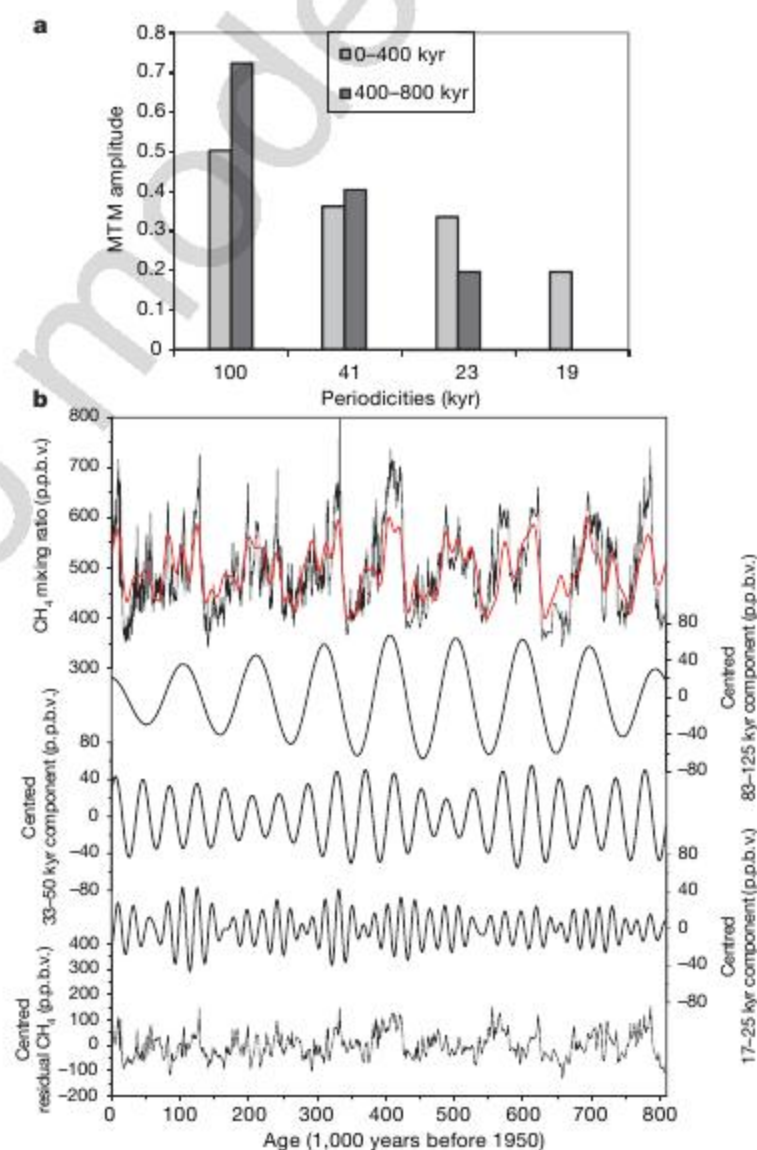
Here, we present the most detailed and longest CH<sub>4</sub> record yet derived from a single ice core: EPICA Dome C (EDC, 75°06'S, 123°20'E, 3,233 m above sea level), reaching back to Marine Isotope Stage (MIS) 20.2 about 802 kyr before present (BP) and covering eight climatic cycles (Fig. 1). We have doubled the EDC CH<sub>4</sub> time resolution (Supplementary Fig. 2) between 0 and 215 kyr BP (MIS 1 to MIS 7, mean resolution: 210 yr) using a melt-refreezing method. EDC samples were measured between 230 and 420 kyr BP (MIS 7 to 11, mean resolution: 390 yr), whereas the record presented in Spahni *et al.*<sup>2</sup> relied on the existing Vostok record<sup>1</sup>. The deepest portion of the EDC core between 3,060.66 m and 3,190.53 m was analysed, providing the first CH<sub>4</sub> data for the time period from 666 to 799 kyr BP (MIS 16 to 20.2, mean resolution: 550 yr; see insert to Fig. 1). The EDC3 ice and gas timescales EDC3 (ref. 5) and EDC3\_gas\_a (ref. 6) are used hereafter, providing good agreement with the composite chronologies<sup>7,8</sup> from EPICA Dronning Maud Land and Greenland (Supplementary Fig. 1) over the last glacial cycle.

With a total of 2,245 individual measurements, the EDC CH<sub>4</sub> record is now complete from the surface down to 3,259.32 m. Because of stratigraphic disturbance in the deepest 60 m (ref. 9), we limit the discussion to the upper 3,190.53 m, which provide an average time resolution of 380 yr over the past 799 kyr. Within analytical uncertainties, our EDC data agree very well with the records of Vostok<sup>1</sup> over the past 420 kyr (Fig. 1) and with the millennial variability recorded in Greenland composite records<sup>7,8</sup> for the last glacial period (Supplementary Fig. 1). The homogeneity, accuracy and resolution of our EDC record make it a new reference for the atmospheric CH<sub>4</sub> temporal evolution on long timescales.

The extended record reveals several new aspects of methane variability. First, the oldest interglacial MIS 19 shows a CH<sub>4</sub> maximum of ~740 p.p.b.v., much higher than during interglacials MIS 13 to 17. The EDC  $\delta D$  record indicates cooler and longer interglacials before 430 kyr, corresponding to the Mid-Brunhes Event transition<sup>9</sup>. This is reinforced by CO<sub>2</sub> (ref. 10) and CH<sub>4</sub> records until MIS 17. But the high CH<sub>4</sub> levels during MIS 19 break down this general trend. Amongst the past nine interglacials, MIS 9 and 19 appear decoupled (with unusually high methane levels) from the overall good correlation ( $r^2 = 0.82$ ) observed between the maximum CH<sub>4</sub> value and the maximum Antarctic warmth<sup>9</sup> (Supplementary Fig. 3). Second, rapid and large fluctuations are identified. The glacial inception from MIS 19 to MIS 18 is accompanied by three distinct CH<sub>4</sub> peaks, having counterparts in the  $\delta D$  (ref. 9) and CO<sub>2</sub> (ref. 10) records. A large CH<sub>4</sub> oscillation (amplitude 170 p.p.b.v.) is identified at the end of MIS 18, comparable in shape with the Younger Dryas event<sup>11</sup>, but with a longer duration (8 kyr between the surrounding maxima; Fig. 1). Its comparison with concomitant CO<sub>2</sub> levels<sup>10</sup> and marine records suggests that it does not reflect a bipolar seesaw sequence of events. A shorter and smaller oscillation (3 kyr between bracketing maxima,

amplitude ~100 p.p.b.v.) is also observed during the early part of interglacial MIS 17.

Wetlands are by far the largest natural source of methane today. The largest wetland extents are found in boreal regions, with a second latitudinal belt between the tropics<sup>12</sup>. Their CH<sub>4</sub> emissions today are shared about one-third boreal and two-thirds tropics. The controversial alternative suggestion of direct plant CH<sub>4</sub> emissions<sup>13</sup> has been recently scaled down substantially<sup>14</sup>. Mean preindustrial CH<sub>4</sub> sources are estimated to be between 200 and 250 Tg yr<sup>-1</sup>, 85% of which comes from wetlands<sup>3</sup>. Temperature changes and water-table variations are the dominant controls of wetland emissions on seasonal and inter-annual timescales<sup>15</sup>. The main sink of methane, oxidation by tropospheric OH radicals, directly feeds back on any CH<sub>4</sub> source change<sup>3</sup>. This is amplified by changes in emissions of volatile organic compounds, which may be related to the reduced extent of forests during glacial conditions<sup>16</sup>. To first order, CH<sub>4</sub> changes during the past 800 kyr should thus reflect varying extent of or emissions from wetlands in different latitudinal belts, and OH feedbacks associated



**Figure 2 | Spectral analysis of the methane record.** **a**, Amplitude of orbital periodicities in the methane record for the time ranges 0–400 kyr and 400–800 kyr, using the multi-taper method (MTM) with normalized data. The 19-kyr component is absent at 400–800 kyr because of its statistical non-significance ( $F$ -test  $< 0.95$ , see Supplementary Information). **b**, Orbital components and residuals (expressed in p.p.b.v. on the centred signal) of the CH<sub>4</sub> record over the past 800 kyr. The red line combines the three periodicities. Note that the amplitude of the orbital components and residuals is similar (~200 p.p.b.v.). The residual CH<sub>4</sub> signal contains other statistically significant periodicities at 4.8 and 8.7 kyr and another periodicity at 10–11 kyr slightly below the confidence limit, not discussed here.

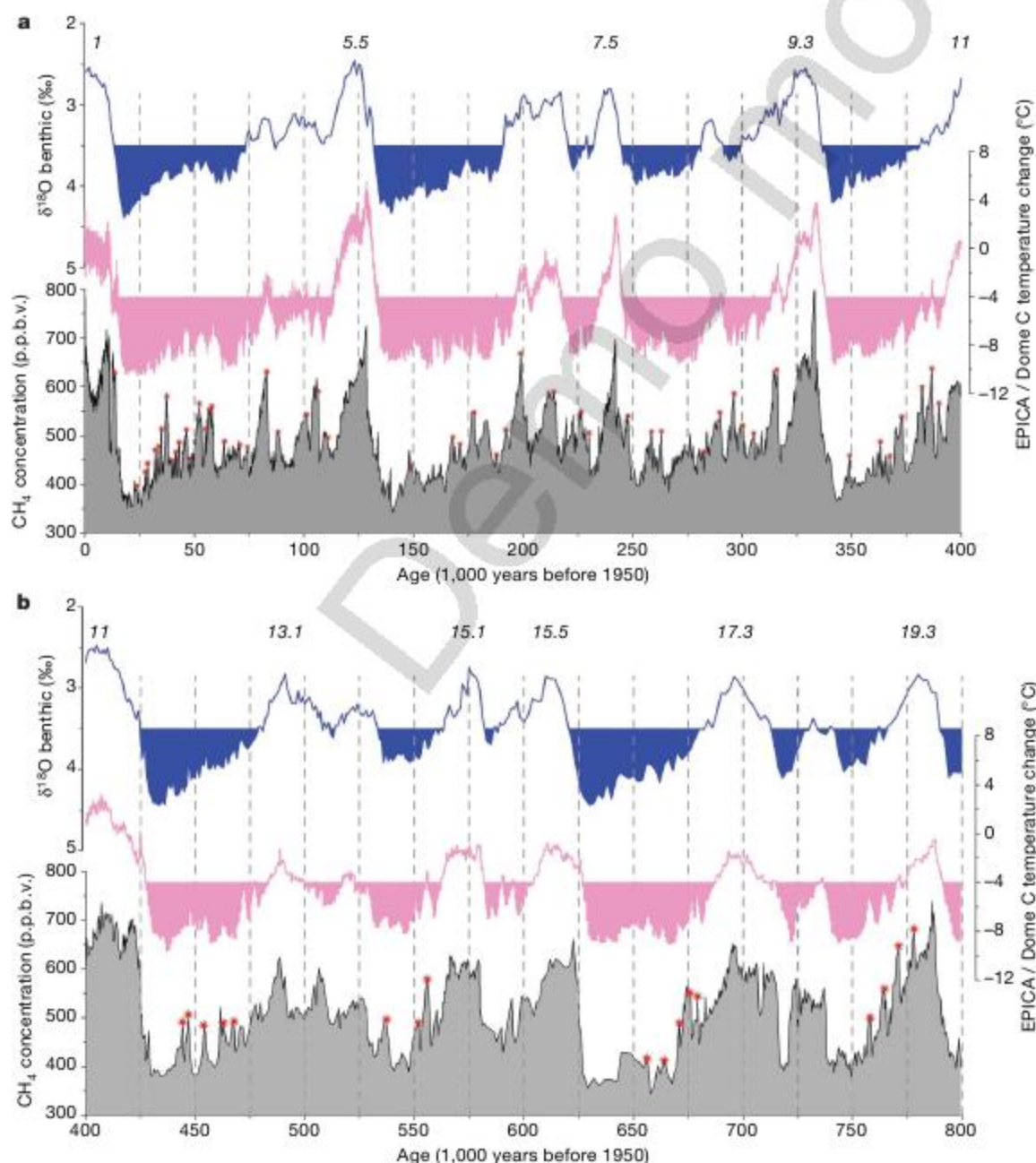


with vegetation changes. Other lesser sources could also have contributed either to long-term trends or to short events, such as biomass burning<sup>17</sup> and clathrate degassing<sup>18</sup>.

The EDC CH<sub>4</sub> spectral analyses (Fig. 2) shed light on the response of methane sources and sinks to orbital forcing. They highlight the orbital periodicities of 100, 41, 23 and 19 kyr already found in the shorter Vostok record<sup>1</sup>. The 100-kyr component dominates the glacial–interglacial variability as in the  $\delta D$  (ref. 9) and global ice volume signals (ICE in Supplementary Fig. 4; deduced for simplification from the stacked oxygen isotopic composition of benthic foraminifera)<sup>19</sup>. But its contribution to the CH<sub>4</sub> variance clearly decreases after the Mid-Brunhes Event, opposed to the trend for  $\delta D$  (ref. 9). This goes together with an increasing contribution of the precessional component towards the present (Fig. 2), reaching the same magnitude as the obliquity component. At high northern latitudes, changes in ice-sheet volume should be the main driver on peat deposition rate, thawing and refreezing of the soil active layer and extent of seasonal snow cover, all affecting CH<sub>4</sub> emissions<sup>20</sup>. A comparison between the EDC CH<sub>4</sub> signal and global ice volume signal<sup>19</sup> does indeed indicate a strong coherency at 100 kyr. The two signals are in phase within current dating uncertainties<sup>5</sup> (Supplementary Fig. 4).

Orbital forcing modulates the latitudinal and land–sea temperature gradients in the Tropics and therefore is a major forcing of

the tropical monsoon systems<sup>21</sup> and of the position of the intertropical convergence zone<sup>22</sup>, which both drive the precipitation rate in the intertropical belt and will influence the areal variability of the large tropical wetlands found in southeast Asia, Africa and South America<sup>12</sup>. Feedbacks from OH associated with varying emissions of volatile organic compounds should closely follow a similar temporal pattern to the regional monsoon evolutions. The tropical climate is dominated mostly by precessional variability<sup>23</sup>, as evidenced for instance by the loess magnetic susceptibility records (loess SUS<sup>24</sup>). Although the entire tropical region affects the methane budget, we compare the CH<sub>4</sub> signal only with the Asian summer monsoon reconstruction from SUS loess, which is currently the only available record of tropical climate variability that covers a time period comparable to our record. The CH<sub>4</sub> and the SUS loess are in phase within dating uncertainties (Supplementary Fig. 4) and show an increasing variance in the precessional band starting around 420 kyr BP, thus suggesting a dominant contribution of monsoon-related processes in the CH<sub>4</sub> variability at precessional periodicities. A picture thus emerges in which atmospheric CH<sub>4</sub> background levels have been modulated by tropical wetlands and/or volatile organic compound emissions from tropical forests during the late Quaternary, with overshoots every 100 kyr associated with varying extents of northern ice sheets and periglacial wetlands. This is in agreement with a recent isotopically constrained CH<sub>4</sub> budget between the last glacial



**Figure 3 | Methane millennial variability.** **a**, 0–400 kyr; **b**, 400–800 kyr. In each panel, from bottom to top, we show the methane signal (grey shaded curve) with 74 identified millennial changes (red stars, mean time between occurrences 6.2 kyr, based on a threshold amplitude of ~50 p.p.b.v. and a correspondence (occurrence and peak-to-peak synchronicity) with an Antarctic isotope maximum<sup>9</sup>. Glacial periods (pink shaded area) are defined by EPICA/Dome C temperature being at least 4 °C below late Holocene values<sup>9</sup>. Time periods of occurrence of millennial variability (blue shaded area) are defined by a threshold value of 3.5‰ in the North Atlantic  $\delta^{18}O$  benthic record<sup>27</sup>. The benthic stack<sup>19</sup> is used to compare with the full CH<sub>4</sub> record over 800 kyr.



maximum and the early Holocene<sup>25</sup>, suggesting a dominating contribution of tropical wetlands in the pre-industrial budget, a switch-on of boreal wetlands when the ice sheets decay and an amplification by the OH sink. Rigorous testing of this picture will require future model simulations of the Earth system that couple climate and the carbon cycle in transient mode over several glacial–interglacial cycles.

On even shorter timescales, high-resolution methane records can be used as a proxy of Northern Hemisphere millennial temperature fluctuations back in time<sup>1</sup>, although the relationship between CH<sub>4</sub> and Greenland temperature is not linear during the last glacial period and the atmospheric CH<sub>4</sub> signal is slightly damped by gas trapping processes<sup>26</sup>. Our new EDC methane record represents a unique continuum of this millennial variability back through the past eight glacial periods with 99% of the record providing a time resolution better than 1,000 yr for the sequence MIS 1–11 (84.5% for MIS 11–20, when additional millennial changes may still be revealed with an improved time resolution). We find 74 millennial CH<sub>4</sub> changes (Fig. 3a and b), defined as an amplitude larger than 50 p.p.b.v. and an association (with peak-to-peak synchronicity) with Antarctic isotope maxima<sup>8,9</sup>. Marine records suggest that the millennial variability shows up whenever the ice-sheet volume, expressed through the oxygen isotopic composition of benthic foraminifera, is above a threshold of 3.5‰ (ref. 27). Antarctic temperature change provides another proxy of glacial conditions, with the threshold considered to be at least 4 °C below late Holocene temperature<sup>9</sup>. Such thresholds, however, fail to capture millennial CH<sub>4</sub> events now identified during the early phase of glacial periods following the warmest interglacials (late MIS 5 and 7; early phase of MIS 8, 10 and 18). The hypothesis of constant lower thresholds is too permissive as it requires CH<sub>4</sub> variability during the lukewarm interglacials (MIS 13 and 17), which is not supported by our data. The combination of CH<sub>4</sub> millennial variability and Antarctic isotope maxima therefore provides a better indicator of bipolar millennial variability, casting doubt on a straightforward link between ice volume or Antarctic cooling and climate instabilities.

## METHODS SUMMARY

The analytical methods of CH<sub>4</sub> measurements in the two laboratories of Bern and LGGE have been already described in detail and compared<sup>2</sup>. The air from polar ice-core samples of about 40 g (Bern) and 50 g (LGGE) is extracted with a melt-refreezing method under vacuum, and the extracted gas is then analysed for CH<sub>4</sub> by gas chromatography. Two standard gases (408 p.p.b.v. CH<sub>4</sub>, 1,050 p.p.b.v. CH<sub>4</sub>) were used at Bern and one (499 p.p.b.v. CH<sub>4</sub>) at LGGE, to calibrate the gas chromatographs. The mean CH<sub>4</sub> analytical uncertainty (1σ) is 10 p.p.b.v. (ref. 2). Concentrations are not corrected for gravitational settling in the firn column as this effect corresponds to only about 1‰ of the measured levels. EPICA/Dome C CH<sub>4</sub> and δD data are on EDC3 age scales<sup>5,6</sup>. The Vostok data were synchronized on the EDC gas age scale by using the program Analyseries<sup>28</sup>. Orbital components and the residual in the CH<sub>4</sub> record are obtained by gaussian filters with the Analyseries program<sup>28</sup>. The precession band is calculated first ( $f = 0.05 \pm 0.01$ ) on the CH<sub>4</sub> signal interpolated every 1 kyr, then the obliquity band on the first residual ( $f = 0.025 \pm 0.005$ ) and then the ~100-kyr band on the second residual ( $f = 0.01 \pm 0.002$ ).

Received 12 October 2007; accepted 17 March 2008.

1. Delmotte, M. *et al.* Atmospheric methane during the last four glacial–interglacial cycles: Rapid changes and their link with Antarctic temperature. *J. Geophys. Res.* **109**, 12104 (2004).
2. Spahni, R. *et al.* Atmospheric methane and nitrous oxide of the Late Pleistocene from Antarctic ice cores. *Science* **310**, 1317–1321 (2005).
3. Denman, K. L. *et al.* in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. *et al.*) 499–587 (Cambridge Univ. Press, 2007).
4. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–111 (2001).
5. Parrenin, F. *et al.* The EDC3 chronology for the EPICA Dome C ice core. *Clim. Past* **3**, 485–497 (2007).

6. Loulergue, L. *et al.* New constraints on the gas age–ice age difference along the EPICA ice cores, 0–50 kyr. *Clim. Past* **3**, 527–540 (2007).
7. Blunier, T. *et al.* Synchronization of ice core records via atmospheric gases. *Clim. Past* **3**, 325–330 (2007).
8. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
9. Jouzel, J. *et al.* Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**, 793–796 (2007).
10. Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* doi:10.1038/nature06949 (this issue).
11. Chappellaz, J. *et al.* Synchronous changes in atmospheric CH<sub>4</sub> and Greenland climate between 40 and 8 kyr BP. *Nature* **366**, 443–445 (1993).
12. Prigent, C., Matthews, E., Aires, F. & Rossow, W. B. Remote sensing of global wetland dynamics with multiple satellite data sets. *Geophys. Res. Lett.* **28**, 4631–4634 (2001).
13. Keppler, F., Hamilton, J. T. G., Brass, M. & Röckmann, T. Methane emissions from terrestrial plants under aerobic conditions. *Nature* **439**, 187–191 (2006).
14. Dueck, T. A. *et al.* No evidence for substantial aerobic methane emission by terrestrial plants; a <sup>13</sup>C-labelling approach. *New Phytol.* **175**, 29–35 (2007).
15. Walter, B. P. & Heimann, M. A process-based, climate-sensitive model to derive methane emissions from natural wetlands: Application to five wetland sites, sensitivity to model parameters, and climate. *Glob. Biogeochem. Cycles* **14**, 745–766 (2000).
16. Kaplan, J. O., Folberth, G. & Hauglustaine, D. A. Role of methane and biogenic volatile organic compound sources in late glacial and Holocene fluctuations of atmospheric methane concentrations. *Glob. Biogeochem. Cycles* **20**, GB2016 (2006).
17. Thonicke, K., Prentice, C. I. & Hewitt, C. D. Modelling glacial–interglacial changes in global fire regimes and trace gas emissions. *Glob. Biogeochem. Cycles* **19**, GB3008 (2005).
18. Kennett, J. P., Cannariato, K. G., Hendy, I. L. & Behl, R. J. Carbon isotopic evidence for methane hydrate instability during quaternary interstadials. *Science* **288**, 128–133 (2000).
19. Lisiecki, L. E. & Raymo, M. E. A Pliocene–Pleistocene stack of 57 globally distributed benthic δ<sup>18</sup>O records. *Paleoceanography* **20**, PA2007 (2005).
20. Schmidt, G. A., Shindell, D. T. & Harder, S. A note on the relationship between ice core methane concentrations and insolation. *Geophys. Res. Lett.* **31**, L23206 (2004).
21. Liu, Z., Otto-Bliesner, B., Kutzbach, J., Li, L. & Shields, C. Coupled climate simulation of the evolution of global monsoons in the Holocene. *J. Clim.* **16**, 2472–2490 (2003).
22. Chiang, J. C. H., Biasutti, M. & Battisti, D. S. Sensitivity of the Atlantic Intertropical Convergence Zone to Last Glacial Maximum boundary conditions. *Paleoceanography* **18**, 1094 (2003).
23. Clement, A. C., Hall, A. & Broccoli, A. J. The importance of precessional signals in the tropical climate. *Clim. Dyn.* **22**, 327–341 (2004).
24. Sun, Y. *et al.* East Asian monsoon variability over the last seven glacial cycles recorded by a loess sequence from the northwestern Chinese Loess Plateau. *Geochim. Geophys. Geosyst.* **7**, Q12Q02 (2006).
25. Fischer, H. *et al.* Changing boreal methane sources and constant biomass burning during the last termination. *Nature* **452**, 864–867 (2008).
26. Spahni, R. *et al.* The attenuation of fast atmospheric CH<sub>4</sub> variations recorded in polar ice cores. *Geophys. Res. Lett.* **30**, 1571 (2003).
27. McManus, J. F. A 0.5-million year record of millennial scale climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
28. Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *EOS Trans. Am. Geophys. Un.* **77**, 379 (1996).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work is a contribution to the European Project for Ice Coring in Antarctica (EPICA), a joint ESF (European Science Foundation)/EC scientific program, funded by the European Commission and by national contributions from Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, Sweden, Switzerland and the United Kingdom. The main logistic support was provided by IPEV and PNRA (at Dome C). Additional funding support of this work was provided by the European FP6 STREP “EPICA-MIS”, by the French ANR PICC (ANR-05-BLAN-0312-01) and by CNRS/INSU programs. We thank the technical team on the field and all those who helped with the methane measurements at Grenoble (B. Bellier, L. Isabelle, E. Estrangin, L. Chan-Tung) and at Bern (G. Hausammann). We also thank P. Yiou and P. Naveau for their spectral analysis courses at LSCE, and H. Fischer and G. Dreyfus for their comments. This is EPICA publication no. 195.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.C. (jerome@lgge.obs.ujf-grenoble.fr).



# Chemical compass model of avian magnetoreception

Kiminori Maeda<sup>1\*</sup>, Kevin B. Henbest<sup>1\*</sup>, Filippo Cintolesi<sup>2</sup>, Ilya Kuprov<sup>2</sup>, Christopher T. Rodgers<sup>2</sup>, Paul A. Liddell<sup>3</sup>, Devens Gust<sup>3</sup>, Christiane R. Timmel<sup>1</sup> & P. J. Hore<sup>2</sup>

Approximately 50 species, including birds, mammals, reptiles, amphibians, fish, crustaceans and insects, are known to use the Earth's magnetic field for orientation and navigation<sup>1</sup>. Birds in particular have been intensively studied, but the biophysical mechanisms that underlie the avian magnetic compass are still poorly understood. One proposal, based on magnetically sensitive free radical reactions<sup>2,3</sup>, is gaining support<sup>4–11</sup> despite the fact that no chemical reaction *in vitro* has been shown to respond to magnetic fields as weak as the Earth's (~50  $\mu$ T) or to be sensitive to the direction of such a field. Here we use spectroscopic observation of a carotenoid–porphyrin–fullerene model system to demonstrate that the lifetime of a photochemically formed radical pair is changed by application of  $\leq 50$   $\mu$ T magnetic fields, and to measure the anisotropic chemical response that is essential for its operation as a chemical compass sensor. These experiments establish the feasibility of chemical magnetoreception and give insight into the structural and dynamic design features required for optimal detection of the direction of the Earth's magnetic field.

Two principal mechanisms for animal magnetoreception have been put forward (reviewed in ref. 1), based on specialized deposits of magnetic iron minerals and magnetically sensitive photochemical reactions, respectively. The latter, the radical pair mechanism, is well established as the source of a variety of magnetic effects on free radical reactions *in vitro*<sup>12,13</sup>. It has been suggested that the avian compass mechanism relies on magneto-sensitive radical pairs formed by photoinduced intramolecular electron transfer reactions in an array of aligned photoreceptors located in the retina<sup>2,3</sup>. A promising candidate radical pair comprises the reduced flavin cofactor and an oxidized tryptophan residue in a cryptochrome flavoprotein<sup>3,8,10</sup>. Such a photochemical process could, in principle, account for two fundamental behavioural characteristics of the avian compass: its dependence on the wavelength of the ambient light<sup>14</sup> and the fact that birds respond to the inclination, rather than the polarity, of the geomagnetic field<sup>15</sup>. This proposal has been corroborated, in part, by the detection of cryptochromes in the retinas of migratory birds<sup>5,6</sup>, and the finding that these proteins are expressed when the birds perform magnetic orientation<sup>6</sup>, and by the observation of light-dependent, cryptochrome-mediated magnetic field effects on plant growth<sup>11</sup>. Theoretical work has also confirmed the principle and clarified some of the details<sup>7–10</sup>. Further compelling evidence for the involvement of radical pairs has come from the observation that weak radiofrequency magnetic fields, which can have profound effects on radical pair reactions *in vitro*<sup>16</sup>, can disrupt the ability of birds to orient in the Earth's magnetic field<sup>4,17</sup>.

Despite numerous studies<sup>12</sup>, it has never been demonstrated that a static magnetic field as weak as that of the Earth can produce detectable changes in chemical reaction rates or product yields. Nor has a radical pair reaction been shown to respond to the direction of such a field, an essential requirement for a compass sensor. For this, it is

essential that at least one of the radicals is immobilized so that its anisotropic magnetic interactions are preserved<sup>2</sup>.

Here we demonstrate, as a proof-of-principle, that a photochemical reaction can act as a magnetic compass. The molecule selected for this purpose is a triad composed of linked carotenoid (C), porphyrin (P) and fullerene (F) groups (Fig. 1)<sup>18</sup>. Green-light irradiation efficiently produces the spin-correlated electronic singlet state of the radical pair (or biradical)  $^1[C^{\bullet+}-P-F^{\bullet-}]$  by sequential intramolecular electron transfers (Fig. 1).  $^1[C^{\bullet+}-P-F^{\bullet-}]$  undergoes reverse electron transfer, either directly to the ground state, with rate constant  $k_S$ , or to the excited triplet state  $^3C-P-F$ , with rate constant  $k_T$ , having first converted to the triplet radical pair,  $^3[C^{\bullet+}-P-F^{\bullet-}]$ . This last process is controlled by the magnetic interactions of the two unpaired electrons and is the magnetic-field-sensitive step. As has been observed for related triads<sup>19,20</sup>, an applied field alters the observed lifetime of  $[C^{\bullet+}-P-F^{\bullet-}]$  by modifying the singlet–triplet character of its spin states, so changing the relative contributions of  $k_S$  and  $k_T$  to the overall kinetics.

We began by characterizing the effects of applied magnetic fields on the disappearance kinetics of the radical pair in isotropic solution. The transient absorption signal of  $C^{\bullet+}$  in  $[C^{\bullet+}-P-F^{\bullet-}]$  at 133 K (Fig. 2a, top), which has a lifetime of ~190 ns in zero field, was markedly increased to ~380 ns in an 8-mT field. The amplitude of the magnetic field effect decreased as the temperature increases and the difference signals were biphasic below ~200 K (Fig. 2a, bottom). Both of these properties are characteristic of a singlet-born radical pair with  $k_T < k_S$ , undergoing spin-lattice relaxation at a rate comparable to its recombination<sup>21</sup>.

The magnetic field dependence of the  $[C^{\bullet+}-P-F^{\bullet-}]$  transient absorption at 119 K (Fig. 2b) shows the biphasic magnetic field response expected for a long-lived radical pair<sup>22</sup>. The change in sign below ~1 mT is the 'low field effect': normally observed for the product yields of radical reactions in solution<sup>12,13</sup>, it was manifested here as a change in the radical pair kinetics. The effect of the applied field on the radical pair absorption was opposite at 100 and 400 ns, as expected from the biphasic time dependence in Fig. 2a (bottom).

Finally, experiments performed in magnetic fields comparable to that of the Earth (Fig. 2c) revealed changes in radical pair absorption of up to ~1.5%. The biphasic time dependence observed at higher fields (Fig. 2a) was inverted here because of the low field effect (Fig. 2b). Thus, for  $t > 400$  ns,  $[C^{\bullet+}-P-F^{\bullet-}]$  recombined more rapidly in a ~50- $\mu$ T field than it did in zero field, which was in turn faster than when the field exceeded 1 mT. This seems to be the first observation of a chemical effect of a magnetic field as weak as ~50  $\mu$ T. The electron Zeeman interaction in such a magnetic field is more than a million times smaller than the thermal energy,  $k_B T$ , implying a negligible effect on the position of a chemical equilibrium or the kinetics of an activated reaction. However, such considerations are irrelevant for the interconversion of singlet and triplet states of radical pairs, a process that is activationless and far from equilibrium.

<sup>1</sup>Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, South Parks Road, Oxford OX1 3QR, UK. <sup>2</sup>Department of Chemistry, University of Oxford, Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK. <sup>3</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA. \*These authors contributed equally to this work.



To function as a chemical compass, a radical pair magnetoreceptor must respond anisotropically to an external magnetic field. We demonstrated this property for  $[C^{\bullet+}-P-F^{\bullet-}]$  using an aligned sample, obtained by freezing C-P-F in the nematic phase of a liquid crystal in the presence of a strong (800 mT) magnetic field. Transient absorption signals,  $A(\theta)$ , were recorded in a 3.1-mT magnetic field, the orientation of which (defined by the angle  $\theta$ ; Fig. 3a) was varied with respect to the alignment axis. The measured  $[C^{\bullet+}-P-F^{\bullet-}]$  absorption (Fig. 3b) shows a clear dependence on  $\theta$  which, similar to the avian magnetic compass, is invariant to inversion of the field direction; that is,  $A(\theta) = A(\theta + \pi)$ .

To obtain additional insight, the anisotropy was also measured by a transient absorption photoselection method (Fig. 3a) using unpolarized laser pump pulses, polarized probe light and a frozen solution of randomly orientated molecules. Calculations (see Supplementary Information) show that the transition dipole moment of the  $C^{\bullet+}$  absorption at 980 nm is almost parallel to the long axis of C-P-F, so that molecules aligned with the polarization axis are preferentially detected. The  $\theta$ -dependence of the  $[C^{\bullet+}-P-F^{\bullet-}]$  absorption,  $A'(\theta)$ , where  $\theta$  is now the angle between the polarization axis and the magnetic-field vector, is essentially identical to that observed in the aligned sample (Fig. 3b, c).

A consideration of the orientational averaging involved in these two experiments (see Supplementary Information) suggests that in both cases, the observed signal should be a weighted sum of even powers of  $\sin\theta$ , but that in the photoselection measurement all terms beyond  $\sin^2\theta$  vanish; that is:

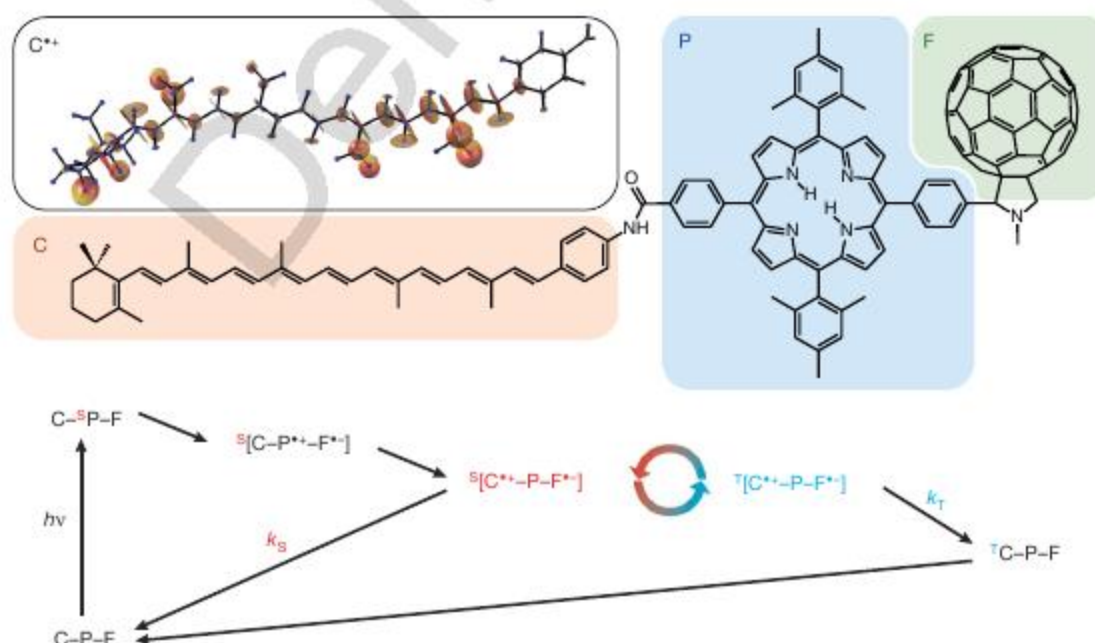
$$A(\theta) \propto \sum_{n=0}^{\infty} a_{2n} \sin^{2n} \theta; \quad A'(\theta) \propto a'_0 + a'_2 \sin^2 \theta$$

Within experimental error, both  $A(\theta)$  and  $A'(\theta)$  have a  $\sin^2\theta$  dependence, consistent with  $a_{2n} \approx 0$  for all  $n > 1$  and a simple orientation dependence of the underlying magnetic field effect (see Supplementary Information).

Our demonstration that a radical pair can act as a chemical compass yields valuable information on the design of a sensitive *in vivo* magnetoreceptor. Although not ideal as a compass,  $[C^{\bullet+}-P-F^{\bullet-}]$  does have several favourable properties—it is rapidly and efficiently

formed by absorption of light and is long enough lived to allow a  $\sim 50 \mu\text{T}$  magnetic field to have a detectable effect on its spin dynamics. Furthermore, the radical centres are sufficiently separated (which minimizes interference from radical–radical exchange and dipolar interactions<sup>22</sup>), whereas their fixed relative orientations and restricted molecular motion in frozen solution result in minimal static and dynamic averaging of the crucial anisotropic magnetic interactions. A further attractive feature of  $[C^{\bullet+}-P-F^{\bullet-}]$  is its highly unsymmetrical distribution of magnetic nuclei (Fig. 1), a feature associated with optimum isotropic and anisotropic magnetic field effects<sup>23</sup>. However, both its recombination and spin relaxation, jointly responsible for the biphasic time dependence (Fig. 2a, c), are fast enough (even at 113 K) to attenuate the  $\sim 50\text{-}\mu\text{T}$  signal (Fig. 2c). Responses closer to the theoretical maximum of 20–40% (ref. 22), expected for spin correlation lifetimes  $\geq 1 \mu\text{s}$  would have allowed the anisotropy measurements (Fig. 3) to be performed at much lower fields than  $\sim 3 \text{ mT}$ . An additional factor responsible for a reduction in the magnetosensitivity of  $[C^{\bullet+}-P-F^{\bullet-}]$  is the extent of delocalization of the unpaired electron in  $C^{\bullet+}$  and the associated cancellation of the anisotropic effects of the many, differently aligned, magnetic hyperfine tensors<sup>8</sup>.

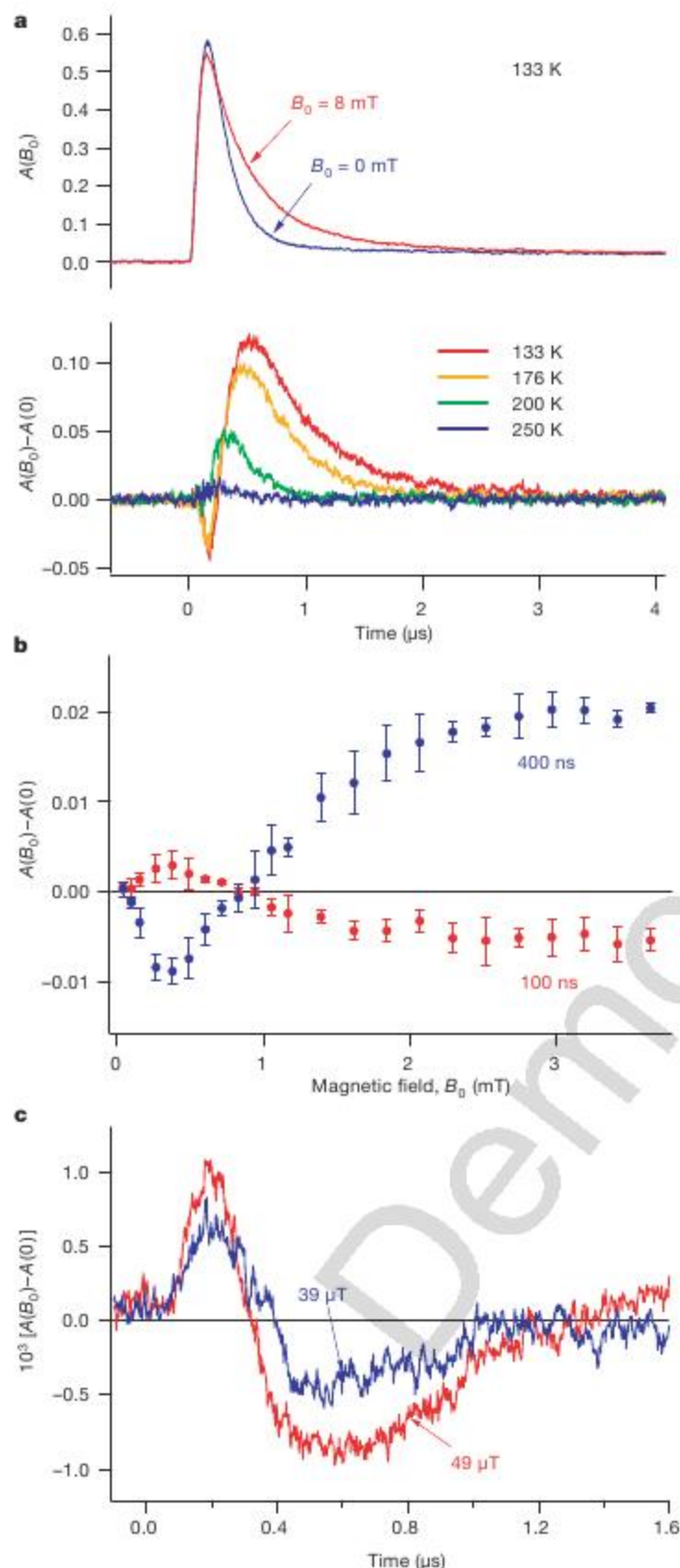
In principle, greater sensitivity to weak magnetic fields would be possible for radical pairs formed in a specialized photoreceptor such as cryptochrome<sup>3,10</sup>. By analogy with photosynthetic charge separation<sup>24</sup>, efficient sequential electron transfer along a chain of tryptophan residues to the flavin cofactor in cryptochrome could produce a well-separated ( $\geq 2 \text{ nm}$ ) radical pair with weak inter-radical interactions<sup>25</sup> within 10 ns (allowing negligible loss of spin correlation), with slow back electron transfer ( $\geq 1 \mu\text{s}$ )<sup>10</sup>. A suitably aligned and immobilized array of such photoreceptors could, at physiological temperatures, have the dynamical properties required for slow spin relaxation ( $\geq 1 \mu\text{s}$ ), a condition that is only approached by  $[C^{\bullet+}-P-F^{\bullet-}]$  at the lowest temperatures studied here. For example, flavin–tyrosine radical pairs that retain their spin correlation for up to 20  $\mu\text{s}$  have been detected in photolyase (a flavoprotein closely related to cryptochrome) at 278 K (ref. 26). In addition, the radicals formed in a photoactive protein may have just a few dominant hyperfine interactions with tensor properties that lead to reinforcement, rather than



**Figure 1 | C-P-F triad.** Structure (top) and reaction scheme (bottom) of the C-P-F triad used to demonstrate the principle of a chemical compass. The interconversion of the singlet (S) and triplet (T) states of the radical pair  $[C^{\bullet+}-P-F^{\bullet-}]$  is driven by magnetic hyperfine interactions and is modulated by the Zeeman interaction with an external magnetic field.  $[C^{\bullet+}-P-F^{\bullet-}]$  can recombine spin-selectively, with rate constants  $k_S$  and  $k_T$ . As the hyperfine

interactions are anisotropic, and  $k_S \neq k_T$ , the lifetime of the radical pair reaction depends on its orientation with respect to the external magnetic field. The inset at top left is a representation of the anisotropic hyperfine interactions in  $C^{\bullet+}$ ; 22 of the 46 protons in this radical have isotropic hyperfine couplings larger than 100  $\mu\text{T}$ .  $F^{\bullet-}$ , by contrast, is almost devoid of hyperfine couplings. *hν*: light excitation.

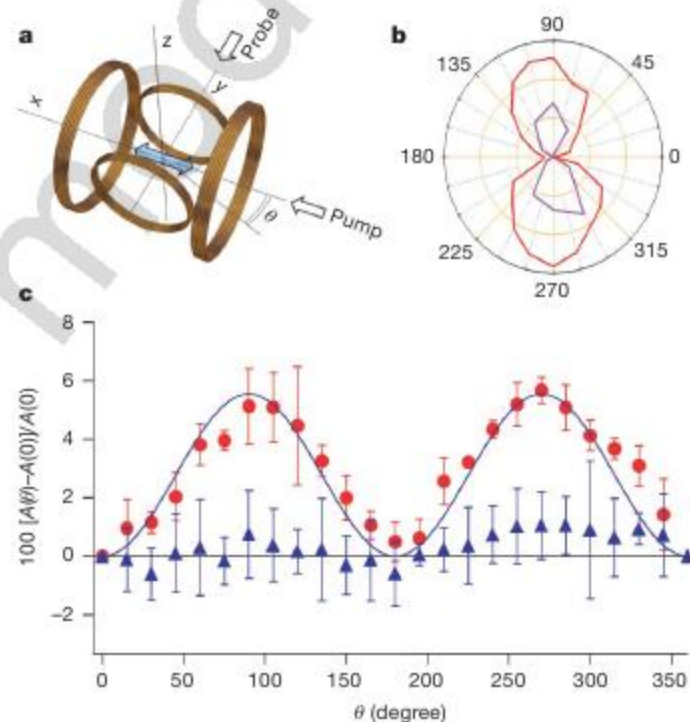




**Figure 2 | Isotropic magnetic field effects on C-P-F.** Transient absorption data recorded for frozen solutions of C-P-F showing the magnetic field-dependence of the recombination of the radical pair state,  $[C^{*+}-P-F^{*-}]$ . Apart from the upper part of **a**, all data are field-on minus field-off differences:  $A(B_0) - A(0)$ . **a**, Top: absorption with and without an 8-mT applied field at 133 K. Bottom: difference signals as a function of temperature. All signals rise with the instrumental time constant ( $\sim 50$  ns). **b**, Changes in the transient absorption of  $[C^{*+}-P-F^{*-}]$  at 119 K averaged over a 50-ns period centred at 100 ns and 400 ns after its formation. **c**, Changes in the transient absorption of  $[C^{*+}-P-F^{*-}]$  at 113 K caused by 39-μT and 49-μT applied magnetic fields. Error bars,  $\pm 1$  s.d.

cancellation, of the anisotropic effects, as seems to be the case for two of the nitrogen atoms in the radical form of the flavin adenine dinucleotide cofactor, FADH $^{\bullet}$  (ref. 8). The favourable asymmetric distribution of hyperfine interactions in  $[C^{*+}-P-F^{*-}]$  might also be found in a cryptochrome. For example, reoxidation of the photochemically reduced flavin cofactor in flavoproteins is mediated by molecular oxygen $^{27}$ . Both  $O_2$  and its reduced form, superoxide  $O_2^{\bullet-}$ , are paramagnetic, have no hyperfine interactions, and if paired with FADH $^{\bullet}$  might allow the reoxidation kinetics to be magnetic field dependent, although it is not clear whether  $O_2^{\bullet-}$ , in particular, would meet the requirement of slow spin relaxation.

These possibilities could be tested *in vitro* in a variety of ways. Spectroscopic measurements of the kind described here could be performed on isolated cryptochromes subject to applied magnetic fields. Radiofrequency magnetic fields could also be used as a diagnostic test for the formation of transient radical pairs and might allow their identification $^{28}$ . The involvement of  $O_2$  could be established by spin-trapping of  $O_2^{\bullet-}$  or by a  $^{17}O$  magnetic isotope effect $^{29}$ .



**Figure 3 | Operation of C-P-F as a chemical compass.** **a**, Schematic of the experimental arrangement used to measure the anisotropy of the magnetic field effect on the disappearance kinetics of  $[C^{*+}-P-F^{*-}]$ . The direction of the applied magnetic field generated by two orthogonal pairs of Helmholtz coils is varied in  $15^\circ$  steps in the  $x$ - $y$  (horizontal) plane. The pump laser pulses and continuous probe light propagate along the  $x$  and  $y$  axes, respectively. The central blue arrow on the  $x$  axis represents the direction of the alignment axis in the experiments on C-P-F dissolved in a frozen nematic liquid crystal, and the probe light polarization axis in the photoselection measurements.  $\theta$  is the angle between the magnetic field vector and the  $x$  axis. **b**, Polar plot of the anisotropy of the magnetic field effect,  $[A(\theta) - A(0)]/A(0)$  (as a function of  $\theta$ , in degrees), on the transient absorption of  $[C^{*+}-P-F^{*-}]$  detected using an aligned sample (purple, 3.1 mT, 193 K) and by photoselection (red, 3.4 mT, 88 K). The maximum magnetic field effects in the two cases were  $\sim 1.5\%$  and  $\sim 5\%$ , respectively. The data for the aligned sample have been doubled for clarity. The anisotropy is smaller in the liquid crystal measurement than in the photoselection experiment mainly because of the faster spin relaxation at the higher temperature of the former. **c**, Data from the photoselection measurements. The red dots show the dependence of the  $[C^{*+}-P-F^{*-}]$  absorption on the direction of the magnetic field,  $\theta$ . The solid line is the best fit to a  $\sin^2 \theta$  form. Also shown (blue) are the signals detected when the polarization axis of the probe light was  $z$  (that is, vertical). No  $\theta$ -dependence is expected or seen. Error bars,  $\pm 1$  s.d.



## METHODS SUMMARY

The C–P–F triad was synthesized according to previously described procedures<sup>18</sup>. Isotropic magnetic field effects were measured using 100  $\mu$ M frozen solutions of C–P–F in 2-methyl-tetrahydrofuran. C–P–F was excited at 532 nm with 7-ns, 5-mJ, 10-Hz repetition rate pulses from a frequency-doubled Nd:YAG laser and the radical pair was detected by means of the C<sup>•+</sup> absorption band at 950 nm using light from a xenon arc lamp. Anisotropic magnetic field effects were measured on an aligned sample and on a frozen isotropic sample using photoselection. For the aligned sample, the conditions were C–P–F (100  $\mu$ M) aligned in E7 liquid crystal (Merck), with the alignment axis parallel to the *x* axis (Fig. 3a). The magnetic field strength was 3.1 mT. The C<sup>•+</sup> absorption was monitored at 900 nm with unpolarized light (xenon arc lamp) and averaged over the period 400  $\pm$  200 ns. Measurements were performed at 193 K. For the frozen isotropic sample, the Nd:YAG pump pulses (*x* direction) were depolarized. The probe light (*y* direction) was supplied by a diode laser ( $\lambda$  = 980 nm, 100 mW), and was polarized in the *x* or *z* direction (see Fig. 3a). The magnetic field strength was 3.4 mT. Measurements were performed at 88 K. For the liquid crystal and photoselection experiments, the magnetic field was generated by means of two sets of orthogonal Helmholtz coils aligned with the *x* and *y* directions (Fig. 3a). They had the same number of turns but different diameters (125 mm and 85 mm, respectively) and carried different currents.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 October 2007; accepted 11 February 2008.  
Published online 30 April 2008.

- Johnsen, S. & Lohmann, K. J. The physics and neurobiology of magnetoreception. *Nature Rev. Neurosci.* **6**, 703–712 (2005).
- Schulten, K., Swenberg, C. E. & Weller, A. A biomagnetic sensory mechanism based on magnetic field modulated coherent electron spin motion. *Z. Phys. Chem. NF111*, 1–5 (1978).
- Ritz, T., Adem, S. & Schulten, K. A model for photoreceptor-based magnetoreception in birds. *Biophys. J.* **78**, 707–718 (2000).
- Ritz, T., Thalau, P., Phillips, J. B., Wiltschko, R. & Wiltschko, W. Resonance effects indicate a radical-pair mechanism for avian magnetic compass. *Nature* **429**, 177–180 (2004).
- Möller, A., Sagasser, S., Wiltschko, W. & Schierwater, B. Retinal cryptochrome in a migratory passerine bird: A possible transducer for the avian magnetic compass. *Naturwissenschaften* **91**, 585–588 (2004).
- Mouritsen, H. et al. Cryptochromes and neuronal-activity markers colocalize in the retina of migratory birds during magnetic orientation. *Proc. Natl Acad. Sci. USA* **101**, 14294–14299 (2004).
- Weaver, J. C., Vaughan, T. E. & Astumian, R. D. Biological sensing of small field differences by magnetically sensitive chemical reactions. *Nature* **405**, 707–709 (2000).
- Cintolesi, F., Ritz, T., Kay, C. W. M., Timmel, C. R. & Hore, P. J. Anisotropic recombination of an immobilized photoinduced radical pair in a 50- $\mu$ T magnetic field: A model avian photomagnetoreceptor. *Chem. Phys.* **294**, 385–399 (2003).
- Wang, K., Mattern, E. & Ritz, T. On the use of magnets to disrupt the physiological compass of birds. *Phys. Biol.* **3**, 220–231 (2006).
- Solov'yov, I. A., Chandler, D. E. & Schulten, K. Magnetic field effects in *Arabidopsis thaliana* cryptochrome-1. *Biophys. J.* **92**, 2711–2726 (2007).
- Ahmad, M., Galland, P., Ritz, T., Wiltschko, R. & Wiltschko, W. Magnetic intensity affects cryptochrome-dependent responses in *Arabidopsis thaliana*. *Planta* **225**, 615–624 (2007).
- Brocklehurst, B. Magnetic fields and radical reactions: Recent developments and their role in nature. *Chem. Soc. Rev.* **31**, 301–311 (2002).
- Timmel, C. R. & Henbest, K. B. A study of spin chemistry in weak magnetic fields. *Phil. Trans. R. Soc. Lond. A* **362**, 2573–2589 (2004).
- Wiltschko, W. & Wiltschko, R. Light-dependent magnetoreception in birds: The behaviour of European robins, *Erithacus rubecula*, under monochromatic light of various wavelengths and intensities. *J. Exp. Biol.* **204**, 3295–3302 (2001).

- Wiltschko, W. & Wiltschko, R. Magnetic compass of European robins. *Science* **176**, 62–64 (1972).
- Rodgers, C. T., Henbest, K. B., Kukura, P., Timmel, C. R. & Hore, P. J. Low-field optically detected EPR spectroscopy of transient photoinduced radical pairs. *J. Phys. Chem. A* **109**, 5035–5041 (2005).
- Thalau, P., Ritz, T., Stapput, K., Wiltschko, R. & Wiltschko, W. Magnetic compass orientation of migratory birds in the presence of a 1.315 MHz oscillating field. *Naturwissenschaften* **92**, 86–90 (2005).
- Kodis, G., Liddell, P. A., Moore, A. L., Moore, T. A. & Gust, D. Synthesis and photochemistry of a carotene-porphyrin-fullerene model photosynthetic reaction center. *J. Phys. Org. Chem.* **17**, 724–734 (2004).
- Kuciauskas, D., Liddell, P. A., Moore, A. L., Moore, T. A. & Gust, D. Magnetic switching of charge separation lifetimes in artificial photosynthetic reaction centers. *J. Am. Chem. Soc.* **120**, 10880–10886 (1998).
- Liddell, P. A. et al. Photoinduced charge separation and charge recombination to a triplet state in a carotene-porphyrin-fullerene triad. *J. Am. Chem. Soc.* **119**, 1400–1405 (1997).
- van Dijk, B., Carpenter, J. K. H., Hoff, A. J. & Hore, P. J. Magnetic field effects on the recombination kinetics of radical pairs. *J. Phys. Chem. B* **102**, 464–472 (1998).
- Timmel, C. R., Till, U., Brocklehurst, B., McLauchlan, K. A. & Hore, P. J. Effects of weak magnetic fields on free radical recombination reactions. *Mol. Phys.* **95**, 71–89 (1998).
- Rodgers, C. T., Norman, S. A., Henbest, K. B., Timmel, C. R. & Hore, P. J. Determination of radical re-encounter probability distributions from magnetic field effects on reaction yields. *J. Am. Chem. Soc.* **129**, 6746–6755 (2007).
- Poluektov, O. G., Paschenko, S. V., Utschig, L. M., Lakshmi, K. V. & Thurnauer, M. C. Bidirectional electron transfer in Photosystem I: Direct evidence from high-frequency time-resolved EPR spectroscopy. *J. Am. Chem. Soc.* **127**, 11910–11911 (2005).
- Efimova, O. E. & Hore, P. J. The role of exchange and dipolar interactions in the radical pair model of the avian magnetic compass. *Biophys. J.* **94**, 1565–1574 (2008).
- Weber, S. Light-driven enzymatic catalysis of DNA repair: A review of recent biophysical studies on photolyase. *Biochim. Biophys. Acta* **1707**, 1–23 (2005).
- Prabhakar, R., Siegbahn, P. E. M., Minaev, B. F. & Agren, H. Activation of triplet dioxygen by glucose oxidase: spin-orbit coupling in the superoxide ion. *J. Phys. Chem. B* **106**, 3742–3750 (2002).
- Henbest, K. B., Kukura, P., Rodgers, C. T., Hore, P. J. & Timmel, C. R. Radio frequency magnetic field effects on a radical recombination reaction: A diagnostic test for the radical pair mechanism. *J. Am. Chem. Soc.* **126**, 8102–8103 (2004).
- Buchachenko, A. L., Kouznetsov, D. A., Orlova, M. A. & Markarian, A. A. Magnetic isotope effect of magnesium in phosphoglycerate kinase phosphorylation. *Proc. Natl Acad. Sci. USA* **102**, 10793–10796 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Ahmad, D. Carbonera, M. di Valentin, G. Giacometti, C. W. M. Kay, P. Raynes, T. Ritz and R. Wiltschko for discussions; N. Baker for technical assistance; and the Oxford Supercomputing Centre for allocation of CPU time. P.J.H., C.R.T. and co-workers are funded by the Engineering and Physical Sciences Research Council, the Human Frontier Science Program, the EMF Biological Research Trust and the Royal Society. D.G. and co-workers are funded by the US National Science Foundation. I.K. is a Fellow by Examination at Magdalen College, Oxford.

**Author Contributions** K.M., K.B.H. and F.C. performed the experiments. K.M., K.B.H. and C.R.T. analysed the data. P.A.L. and D.G. synthesized the triad molecule. C.T.R. and P.J.H. analysed the orientational averaging. I.K. performed *ab initio* calculations. F.C., C.R.T. and P.J.H. designed the study. C.R.T. co-ordinated the study. P.J.H. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.J.H. (peter.hore@chem.ox.ac.uk) or C.R.T. (christiane.timmel@chem.ox.ac.uk).



## METHODS

**Isotropic magnetic field effects.** Transient absorption signals of  $C^{*+}$  in 2-methyl-tetrahydrofuran (MTHF) were detected with an Oriel 77250 Series 1/8 m monochromator, a grating optimized for maximum reflectivity at 1,000 nm and a Hamamatsu photomultiplier R5108. The magnetic field was produced by a pair of Helmholtz coils (radius 125 mm). Alternate measurements with and without the applied magnetic field were performed in synchrony with the laser. The temperature of the sample, held in a cryostat, was controlled with flowing nitrogen gas. The cryostat and field coils were enclosed in a mu-metal box to shield the sample from the Earth's magnetic field. The field inside the box was  $0 \pm 10 \mu\text{T}$  (measured with high-sensitivity one- and three-dimensional Hall transducers).

**Anisotropic magnetic field effects.** C-P-F (100  $\mu\text{M}$ ) was dissolved in E7 liquid crystal (Merck) and aligned in a 800-mT magnetic field by cycling the temperature three times between 20 °C and 60 °C (between the nematic and isotropic phases, respectively). The aligned sample was allowed to cool in the magnetic field and was then transferred into the cryostat and cooled further to 193 K. The solution was sufficiently viscous that the cooling and subsequent transfer could be done without loss of alignment.

For the photoselection experiments, a Halbo Optics WDQ15M depolarizer was used to depolarize the Nd:YAG pump pulses. A Thorlabs L980P100 diode laser supplied the probe light, which was polarized using a Glan Thompson polarizer.

Transient absorption data were measured for each of 24 equally spaced values of  $\theta$  (Fig. 3a) in random order and sorted according to  $\theta$  in the computer memory. This process was repeated 300 times for the photoselection experiments or 1,500 times for the liquid crystal sample with a different sampling order each time, and the resulting signals averaged for each  $\theta$ . In all other respects, the experiments were performed essentially as described for the MTHF solutions.

To confirm the anisotropic behaviour of the magnetic field effect, the experiment was repeated with a 2.6-mT field in the  $x$ - $z$  plane, and the probe light still directed along the  $y$  axis. When the polarization axis of the probe light was rotated by  $\pm 45^\circ$  in the  $x$ - $z$  plane, the absorption was phase-shifted by  $\pm 45^\circ$  (Supplementary Fig. 1).

**Magnetic fields.** Calibration measurements on the two sets of orthogonal Helmholtz coils used for the liquid crystal and photoselection experiments revealed that a 46% larger current was required in the  $x$  coils to achieve the same magnetic field strength as the  $y$  coils. The currents through the two coils were:

$$I_x(\theta_n) = 1.46 I_0 \cos \theta_n$$

$$I_y(\theta_n) = I_0 \sin \theta_n$$

where  $\theta_n = n\delta$ ,  $\delta = 15^\circ$  and  $n = 0, 1, \dots, 23$ .

The magnetic fields produced by the two sets of coils were measured with a 3D Hall probe and are shown as the dashed lines in Supplementary Fig. 2. The black dashed line is  $\sqrt{I_x^2(\theta_n) + I_y^2(\theta_n)}$ . The slight variation in this quantity with  $\theta_n$  was reduced by modifying the currents using a table of correction factors  $g_n$ :

$$I_x = 1.46 I_0 \cos \theta_n / g_n$$

$$I_y = I_0 \sin \theta_n / g_n$$

The values of  $g_n$  were in the range 0.996–1.027. The resulting current (Supplementary Fig. 2, solid black line) showed a reduced variation with  $\theta$ . The averaged magnetic field for the photoselection experiments was 3.4 mT, with a standard deviation of 0.02 mT.



# Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4*

Marc Hanikenne<sup>1†\*</sup>, Ina N. Talke<sup>1†\*</sup>, Michael J. Haydon<sup>1†</sup>, Christa Lanz<sup>2</sup>, Andrea Nolte<sup>1†</sup>, Patrick Motte<sup>3,4</sup>, Juergen Kroymann<sup>5</sup>, Detlef Weigel<sup>2</sup> & Ute Krämer<sup>1†</sup>

Little is known about the types of mutations underlying the evolution of species-specific traits. The metal hyperaccumulator *Arabidopsis halleri* has the rare ability to colonize heavy-metal-polluted soils, and, as an extremophile sister species of *Arabidopsis thaliana*, it is a powerful model for research on adaptation<sup>1–3</sup>. *A. halleri* naturally accumulates and tolerates leaf concentrations as high as 2.2% zinc and 0.28% cadmium in dry biomass<sup>4</sup>. On the basis of transcriptomics studies, metal hyperaccumulation in *A. halleri* has been associated with more than 30 candidate genes that are expressed at higher levels in *A. halleri* than in *A. thaliana*<sup>4–6</sup>. Some of these genes have been genetically mapped to broad chromosomal segments of between 4 and 24 cM co-segregating with Zn and Cd hypertolerance<sup>7–9</sup>. However, the *in planta* loss-of-function approaches required to demonstrate the contribution of a given candidate gene to metal hyperaccumulation or hypertolerance have not been pursued to date. Using RNA interference to downregulate *HMA4* (*HEAVY METAL ATPASE 4*) expression, we show here that Zn hyperaccumulation and full hypertolerance to Cd and Zn in *A. halleri* depend on the metal pump *HMA4*. Contrary to a postulated global *trans* regulatory factor governing high expression of numerous metal hyperaccumulation genes, we demonstrate that enhanced expression of *HMA4* in *A. halleri* is attributable to a combination of modified *cis*-regulatory sequences and copy number expansion, in comparison to *A. thaliana*. Transfer of an *A. halleri* *HMA4* gene to *A. thaliana* recapitulates Zn partitioning into xylem vessels and the constitutive transcriptional upregulation of Zn deficiency response genes characteristic of Zn hyperaccumulators. Our results demonstrate the importance of *cis*-regulatory mutations and gene copy number expansion in the evolution of a complex naturally selected extreme trait<sup>10</sup>. The elucidation of a natural strategy for metal hyperaccumulation enables the rational design of technologies for the clean-up of metal-contaminated soils and for bio-fortification.

*HMA4* (ref. 4) is among a large number of genes more highly expressed in *A. halleri* than in *A. thaliana*<sup>3,6</sup> and encodes a plasma membrane protein of the *P<sub>1B</sub>* family of transition metal pumps in the P-type ATPase superfamily<sup>8,11</sup>. To investigate whether *A. halleri* *HMA4* (*AhHMA4*) functions in metal hyperaccumulation or hypertolerance of *A. halleri*, we reduced the expression of *AhHMA4* by RNA interference (RNAi)<sup>12</sup> (Supplementary Fig. 1). For this, we developed a genetic transformation system, which has not so far been available for hyperaccumulator species. *HMA4* transcript levels were decreased to between 45% and 10% of wild-type levels in different

RNAi lines (Fig. 1a, Supplementary Fig. 2), which appeared morphologically normal. After cultivation of *HMA4* RNAi lines in hydroponic solutions, their shoots contained only 12–35% of the Zn concentrations found in wild-type *A. halleri* (Fig. 1b, Supplementary Fig. 3). These levels were similar to shoot Zn concentrations in *A. thaliana* as a representative non-accumulator<sup>4</sup>. In roots of wild-type *A. halleri*, metal concentrations are low, reflecting high root-to-shoot metal fluxes<sup>4,13,14</sup>. By comparison, roots of *HMA4* RNAi lines accumulated 49- to 134-fold higher Zn concentrations, again similar to the non-accumulator *A. thaliana* (Fig. 1c). Partitioning and accumulation of metals other than Zn and Cd (Supplementary Fig. 4) were not consistently changed in *HMA4* RNAi lines. Together these results demonstrate that high *HMA4* transcript levels are required for highly efficient root-to-shoot Zn flux and for Zn hyperaccumulation in the shoots of *A. halleri*. This is distinct from the function of *A. thaliana* *HMA4* (*AtHMA4*) in root-to-shoot translocation of Zn for the maintenance of Zn-dependent processes in the shoot<sup>15,16</sup>.

We next examined how *HMA4* alters the distribution of Zn within the roots of *A. halleri*, by imaging with the fluorescent Zn indicator Zinpyr-1<sup>16</sup>. In the wild type, fluorescence was most intense in the xylem vessels located inwards from the pericycle cell layer of the root vasculature (Fig. 1d, Supplementary Fig. 5). By contrast, in the *HMA4* RNAi lines signal intensity was maximal in the pericycle cell layer, qualitatively similar to the change in Zn localization reported for the *A. thaliana* *hma4* mutant compared to the wild type<sup>15</sup>. This indicates that silencing of *AhHMA4* impairs the release of Zn from the root symplast into the apoplastic xylem vessels, which provide the primary pathway for the movement of solutes from the root to the shoot with the transpiration stream<sup>17</sup>. These results are in agreement with the decreased Zn partitioning into the shoots in *HMA4* RNAi lines.

If increased *HMA4* activity is a primary event in heavy metal hyperaccumulation, we would expect that this affects other candidate genes for hyperaccumulation known to be highly expressed in *A. halleri*<sup>4</sup>. We therefore analysed RNA from plants cultivated in a 5  $\mu$ M Zn solution by real-time RT-PCR. Expression of Zn deficiency response genes<sup>4,18</sup> was specifically decreased in roots of the *HMA4* RNAi lines. For example, relative transcript levels of *IRT3* (Fig. 1e) and *ZIP4* (Fig. 1f) were correlated with *HMA4* transcript levels across all genotypes ( $r = 0.99$  and  $0.98$ , respectively). The membrane transporters encoded by *IRT3* and *ZIP4* are both members of the same family of proteins implicated in the cellular uptake of Zn (ref. 18). Thus, the high expression of these genes in roots of wild-type *A. halleri*<sup>4</sup> appears to be a secondary consequence of increased *HMA4* activity, and is likely to further enhance shoot metal accumulation.

<sup>1</sup>Max Planck Institute of Molecular Plant Physiology, D-14476 Potsdam, Germany. <sup>2</sup>Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany. <sup>3</sup>Plant Cell Biology, Department of Life Sciences, <sup>4</sup>Center for Assistance in Technology of Microscopy, Department of Chemistry, University of Liège, B-4000 Liège, Belgium. <sup>5</sup>Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany. <sup>†</sup>Present addresses: Plant Cell Biology, Department of Life Sciences, University of Liège, B-4000 Liège, Belgium (M.H.); Department of Thoracic, Cardiac and Vascular Surgery, University of Tübingen, D-72076 Tübingen, Germany (A.N.); BIOQUANT Center, Heidelberg Institute of Plant Sciences, University of Heidelberg, D-69120 Heidelberg, Germany (I.N.T., M.J.H., U.K.).

\*These authors contributed equally to this work.

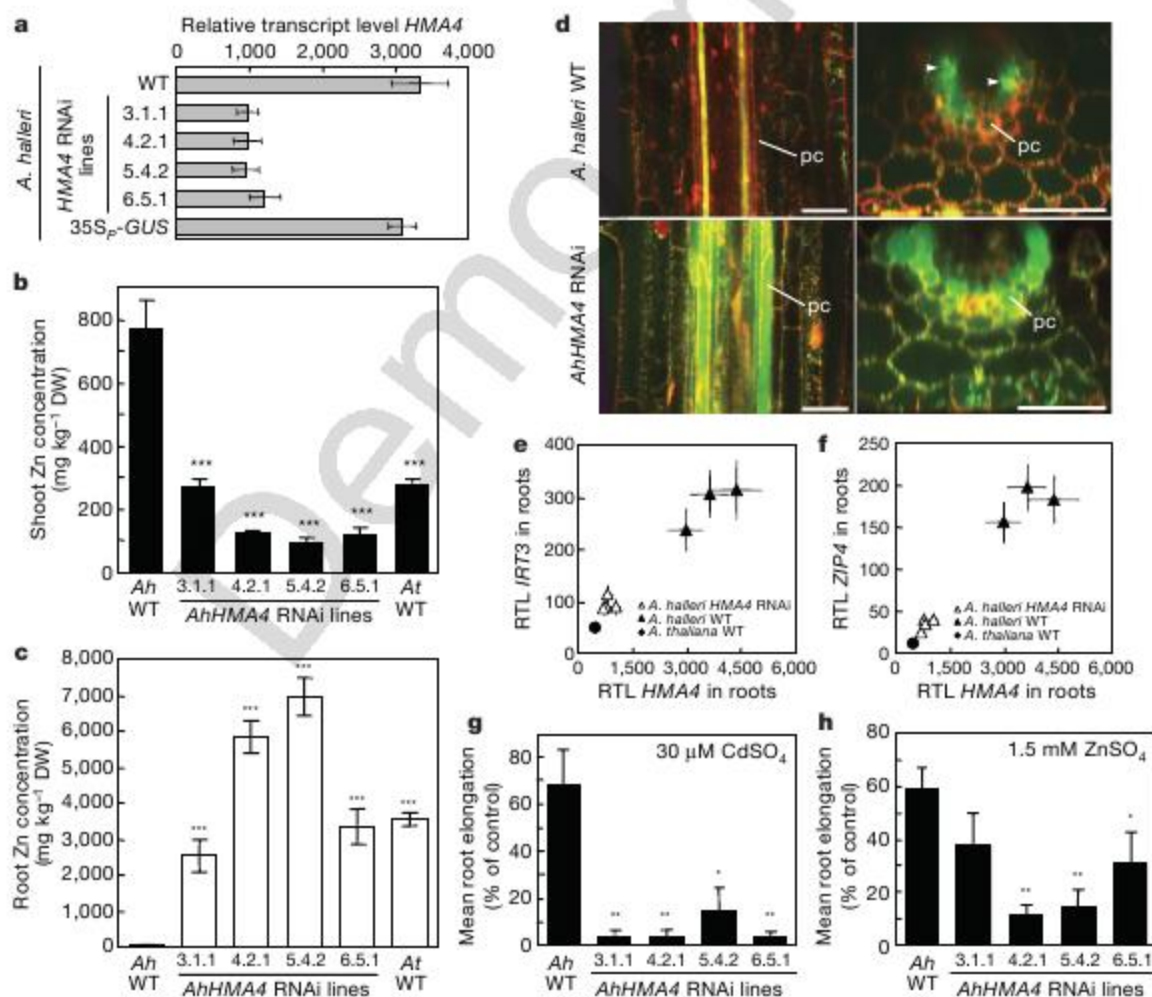


Heterologous expression of *AhHMA4* in cells of the budding yeast *Saccharomyces cerevisiae* was reported to confer metal tolerance<sup>4</sup> and metal hypersensitivity<sup>8</sup>. Recently, three respective major QTL (quantitative trait loci) for Zn and Cd hypertolerance were mapped in progeny from a cross between *A. halleri* and its metal-sensitive non-accumulator relative *Arabidopsis lyrata*<sup>8,9</sup>. Notably, major QTL for both Zn and Cd hypertolerance contained *HMA4*, but comprised several centimorgans and thus at least several hundred genes. To examine whether *HMA4* functions not only in metal hyperaccumulation of *A. halleri*, but also in hypertolerance, we determined the effects of toxic metal concentrations on root growth of plants in hydroponic solutions<sup>19,20</sup>. In the presence of 30  $\mu$ M Cd or 1.5 mM Zn, root elongation in wild-type *A. halleri* plants was still about 68% and 60% of the elongation under control conditions, respectively. In contrast, roots of *HMA4* RNAi lines exhibited only 3–15% and 12–37% of the elongation under control conditions, respectively (Fig. 1g, h). These data demonstrate that *HMA4* makes a substantial contribution to Cd and Zn hypertolerance in *A. halleri*.

Our results indicate that *HMA4* has a key role in several aspects of the complex metal hyperaccumulator phenotype. We therefore addressed the molecular basis for elevated *HMA4* expression in *A. halleri* compared to the non-accumulator *A. thaliana*. Earlier, Southern blot screening provided circumstantial evidence that there may be more than a single *HMA4* gene copy in the genome of *A. halleri*<sup>4</sup>. However, information on the number of gene copies, their respective sequences, expression and functionality has been unavailable. To examine this, two overlapping *A. halleri* genomic BAC

(bacterial artificial chromosome) clones were sequenced. Compared to the syntenic segment in *A. thaliana*, there is a complex triplication of a region that includes *HMA4* (corresponding to At2g19110 in *A. thaliana*) and the two downstream genes (At2g19120 and At2g19130) (Fig. 2a). In addition to partial deletions and inversions of At2g19120 and At2g19130 orthologues, several transposon and retrotransposon insertions are evident. The flanking sequences are largely syntenic with *A. thaliana*. Within coding sequences, the three *AhHMA4* gene copies are on average 99% identical to each other, but share only 88% sequence identity with *AtHMA4*, suggesting that the triplication may have occurred relatively recently in the *A. halleri* lineage (Supplementary Table 1). After reverse transcription we specifically quantified either the *HMA4-1* complementary DNA alone or both the almost identical *HMA4-2* and *HMA4-3* cDNAs by real-time PCR (Supplementary Fig. 6). The results are consistent with similar expression levels of all three *HMA4* gene copies (Fig. 2b). Thus, both an increase in copy number and elevated expression of individual gene copies contribute to the high transcript levels of *HMA4* in *A. halleri* in comparison with *A. thaliana*.

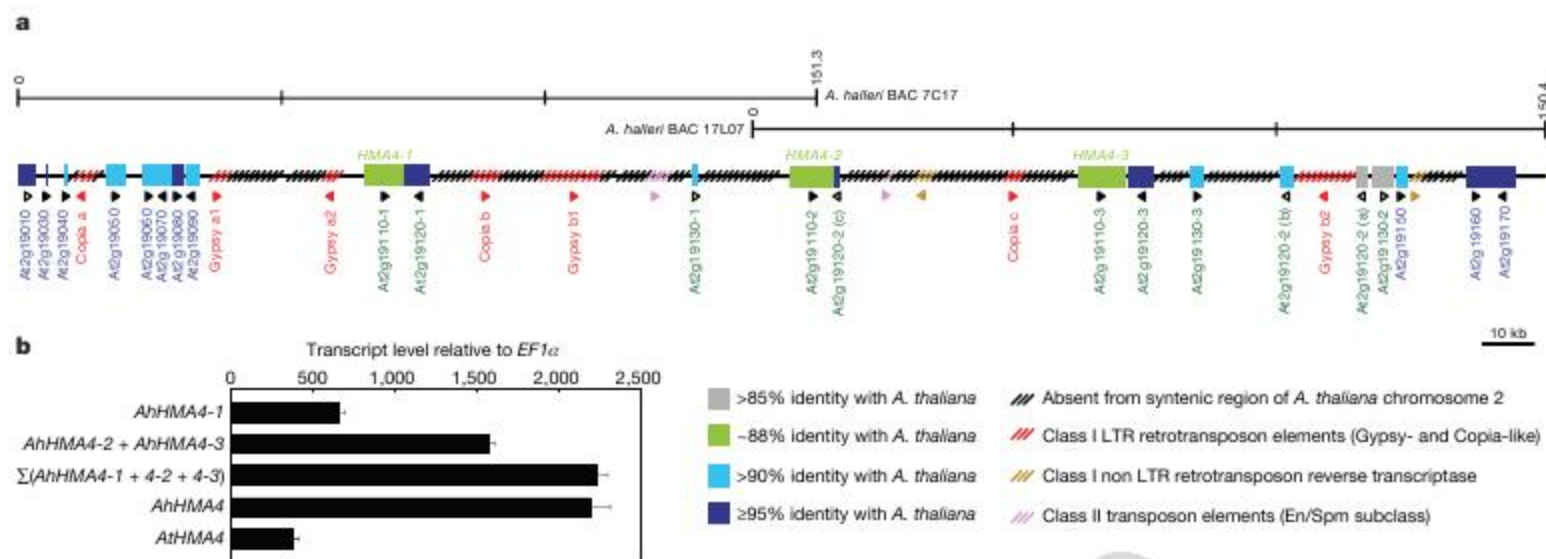
Sequence divergence from *A. thaliana* is most pronounced in the 5'-flanking regions of *AhHMA4* genes (Supplementary Table 1)<sup>21</sup>, suggesting possible regulatory innovations in *A. halleri*. To determine experimentally whether the increased expression of individual *HMA4* gene copies is due to *cis*-regulatory changes, we generated fusions of the promoters of *AtHMA4* and of the three *AhHMA4* gene copies to the  $\beta$ -glucuronidase (*GUS*) reporter for the transformation of both *A. halleri* and *A. thaliana*. Reporter activity directed by the



**Figure 1 | Characterization of *A. halleri* *HMA4* RNAi lines.** **a**, *AhHMA4* transcript levels relative to *EF1 $\alpha$*  in roots of *A. halleri* wild type (WT) and 35S<sub>p</sub>-*GUS* transformant controls, and four *HMA4* RNAi lines. **b**, **c**, Zinc concentrations in shoots (**b**) and roots (**c**) of *A. halleri* (*Ah*) genotypes and *A. thaliana* Col (*At*) WT, cultivated in a hydroponic solution containing 5  $\mu$ M Zn for 3 weeks (DW, dry weight). **d**, Zn localization in *A. halleri* roots. Confocal images show fluorescent Zn signals (green) and cell walls (red).

Scale bars, 50  $\mu$ m; pc, pericycle; arrowheads highlight fluorescent Zn signals in xylem vessels. **e**, **f**, Relative transcript levels (RTL) of *IRT3* (**e**) and *ZIP4* (**f**) plotted against relative *HMA4* transcript levels, all normalized to *EF1 $\alpha$* , in different genotypes. **g**, Cadmium tolerance, and **h**, zinc tolerance of *A. halleri* genotypes. Values are mean  $\pm$  s.e.m.,  $n = 4$  to 6 and 12 to 18 individuals for RNAi lines and wild type, respectively (**b**, **c**, **g**, **h**),  $n = 6$  to 8 (**e**, **f**). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



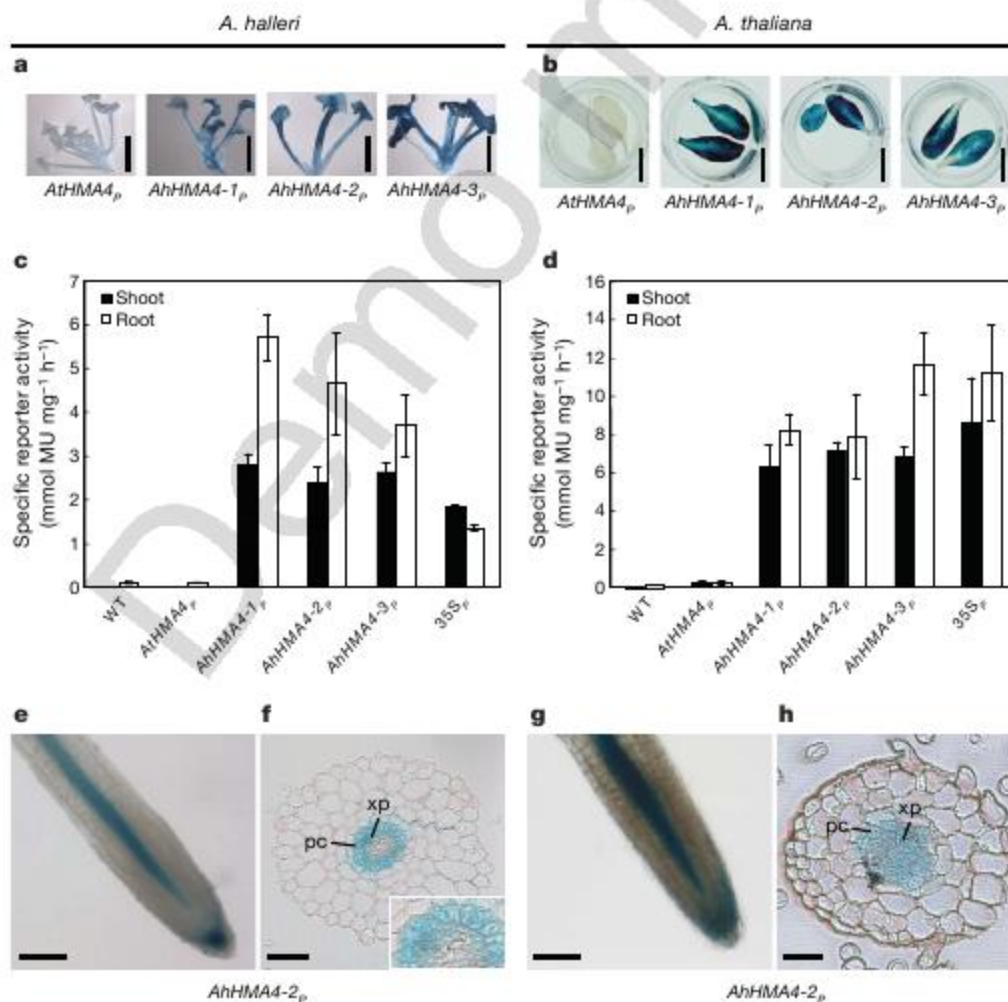


**Figure 2 | Genomic organization and expression of *HMA4* genes in *A. halleri*.** **a**, Diagram showing two overlapping *A. halleri* genomic BACs. Genes are represented by boxes and named by *A. thaliana* Genome Identifier codes (blue text, regions syntenic with *A. thaliana*; green, triplicated genes). Triangles indicate the direction of transcription (filled, intact genes; open,

truncated or pseudogenes). Three fragments of one At2g19120 orthologue are denoted (a), (b) and (c); LTR, long terminal repeat. **b**, Relative transcript levels of *HMA4* gene copies in *A. halleri* roots, and of *AtHMA4* in *A. thaliana* roots for comparison (mean  $\pm$  s.e.m.,  $n = 6$ ).

*AtHMA4* promoter (*AtHMA4<sub>p</sub>*) was much lower than that directed by any of the three *AhHMA4* promoters (*AhHMA4-1<sub>p</sub>*, *AhHMA4-2<sub>p</sub>*, *AhHMA4-3<sub>p</sub>*), which were as effective as the strong constitutive

cauliflower mosaic virus (CaMV) 35S promoter, 35S<sub>p</sub> (Fig. 3a–d). Spatial patterns of reporter activity for all three *AhHMA4<sub>p</sub>* constructs were highly similar in both *A. halleri* and *A. thaliana* (Fig. 3e, g,



**Figure 3 | Levels and cell specificity of *HMA4* promoter activity in *A. halleri* and *A. thaliana*.** **a, c, e, f**, *A. halleri* (left); **b, d, g, h**, *A. thaliana* (right). **a, b**, Histochemical detection of GUS activity (blue), directed by the *A. thaliana* or *A. halleri* *HMA4* promoters, in leaves of 5-week-old plants. **c, d**, Specific GUS activity in protein extracts (mean  $\pm$  s.e.m.,  $n = 4$  to 7 independent lines; MU, 4-methylumbelliferone). Values for the CaMV 35S promoter are shown for comparison. **e–h**, Reporter activity in whole mounts

(**e, g**) and transverse sections (**f, h**) of root tips from *AhHMA4-2<sub>p</sub>*-GUS transformants representative of all *AhHMA4* promoters (Supplementary Figs 7, 8 and 10). Tissues were stained for 18 h (**a, b**) or 0.5 h (**e** to **h**). The inset (**f**) shows a close-up of the vascular cylinder. Scale bars: **a, b**, 5 mm; **e, g**, 100  $\mu$ m; **f, h**, 15  $\mu$ m. pc, pericycle; xp, xylem parenchyma. Root anatomy differs slightly between *A. thaliana* and *A. halleri*.



Supplementary Figs 7–10). Qualitatively, activity patterns were also similar for promoters originating from both species (Supplementary Figs 7, 8)<sup>15</sup>. The *HMA4* expression pattern was confirmed by *in situ* hybridization in wild-type *A. halleri* (Supplementary Figs 9, 10). This revealed specific messenger RNA accumulation in the root pericycle and xylem parenchyma (Fig. 3f, h, Supplementary Fig. 10), supporting a function for *HMA4* in xylem loading of Zn (see Fig. 1d). Expression in the xylem parenchyma and the cambium of leaves (Supplementary Fig. 9) is consistent with a possible role of *HMA4* in metal distribution within the leaf blade and the exclusion of metals from specific cell types. Together, our data confirm that high *HMA4* transcript levels in *A. halleri* are the result of gene copy number triplication combined with the enhanced expression of all three *A. halleri* *HMA4* genes specified in *cis*.

Finally, we determined whether increased *HMA4* activity is not only necessary but also sufficient for altered heavy metal accumulation and tolerance. To this end, we transformed *A. thaliana* with an *AhHMA4* mini-gene, consisting of an *AhHMA4* cDNA linked to the *AhHMA4-1* promoter. Primary transformants contained moderately elevated *HMA4* transcript levels ( $2.44 \pm 0.08$  and  $2.85 \pm 0.89$  fold in roots and shoots, respectively, compared to wild type, mean  $\pm$  s.e.m.,

$n = 6$ ; Supplementary Fig. 9m–o), at the lower end of the range expected based on previous comparisons between *A. halleri* and *A. thaliana*<sup>4</sup>. Zinpyr-1 imaging showed high levels of Zn in xylem tissues of roots of transgenic plants (Fig. 4a), mimicking the distribution of Zn in wild-type *A. halleri* (see Fig. 1d). By contrast, in non-transgenic *A. thaliana* Zn levels were highest in the root pericycle cell layer (Fig. 4a), resembling the localization in *A. halleri* *HMA4* RNAi lines (see Fig. 1d). Thus, transformation of *A. thaliana* with *AhHMA4* is sufficient to recapitulate Zn distribution typical of *A. halleri* in the roots of the non-hyperaccumulator *A. thaliana*. Roots of *AhHMA4*-transformed *A. thaliana* also contained increased transcript levels of the Zn deficiency response genes *ZIP4* and *IRT3* (Fig. 4b), in analogy with wild-type *A. halleri* (see Fig. 1e, f). These results further support the model that in roots of *A. halleri*, the high *HMA4*-dependent Zn flux into the xylem depletes symplastic Zn pools, thereby triggering the upregulation of Zn deficiency response genes in the roots<sup>4</sup>.

*A. thaliana* *AhHMA4* transformants grown in media supplemented with toxic concentrations of 150  $\mu$ M Zn or 40  $\mu$ M Cd developed enhanced leaf chlorosis and smaller rosettes, which are signs of Zn and Cd hypersensitivity of the shoots (Fig. 4c, Supplementary Fig. 11). When cultivated in 5  $\mu$ M Zn, *AhHMA4* transformants were healthy and accumulated slightly higher Zn concentrations than non-transgenic plants ( $1.16 \pm 0.07$  fold, mean  $\pm$  s.e.m.,  $n = 4$  independent experiments,  $P < 0.05$ ; data not shown). These results suggest a more efficient transfer of metals from roots to leaves in *AhHMA4* transformants compared to non-transgenic *A. thaliana*. The metal sensitivity of shoots of *A. thaliana* expressing *AhHMA4* indicates that additional genes are required for metal detoxification in order to accommodate the high *HMA4*-dependent metal flux into the shoots of *A. halleri*<sup>7,9,22</sup>.

In summary, using a series of functional criteria<sup>23</sup>, we have demonstrated a major role for *AhHMA4* in naturally selected Zn hyperaccumulation and associated Cd and Zn hypertolerance in *A. halleri*. High *HMA4* expression in *A. halleri* is specified in *cis* and amplified by gene copy number expansion. Increased expression of *HMA4* in *A. halleri* supports the enhanced Zn flux from the root symplasm into the xylem vessels necessary for shoot Zn hyperaccumulation, and acts as a physiological master switch to upregulate Zn deficiency response gene expression in roots (Supplementary Fig. 1). These findings are of central importance for the development of phytoremediation and bio-fortification strategies. Taken together, our results provide evidence for previously proposed roles of *cis* regulatory diversification<sup>24</sup> and copy number expansion<sup>25,26</sup> in eukaryotic adaptation<sup>27</sup>.

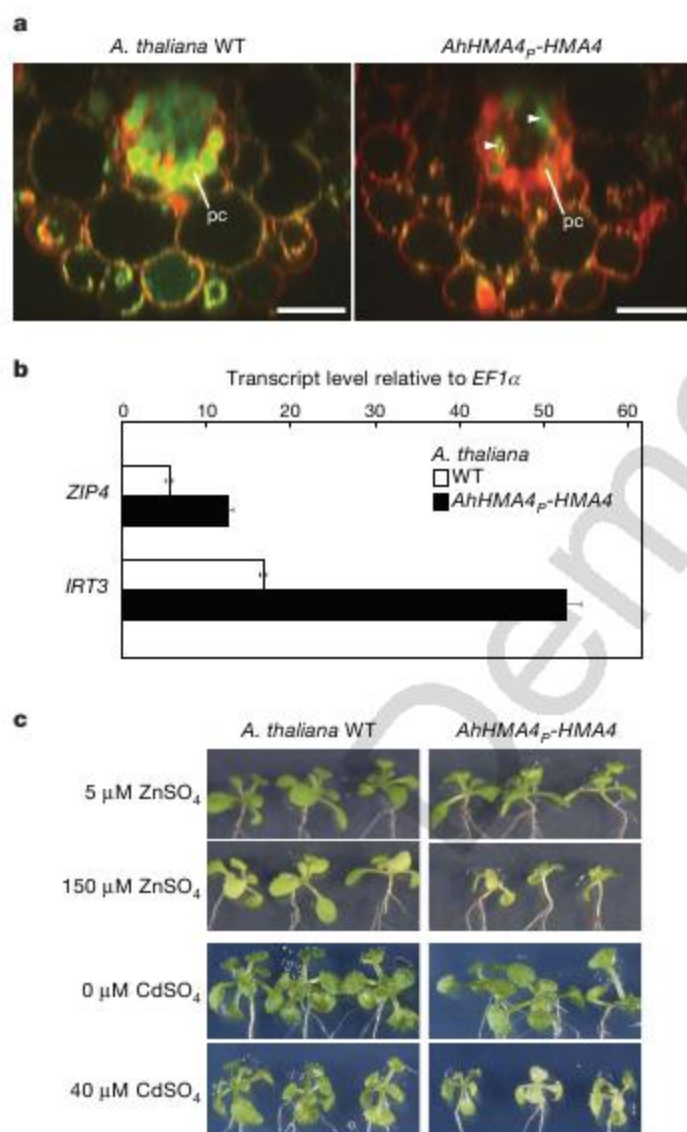
## METHODS SUMMARY

**Plant material.** *A. halleri* (L.) O'Kane and Al-Shehbaz ssp. *halleri* (accession Langelsheim) or *A. thaliana* L. Heynhold (accession Columbia) were used in all experiments. Plants were cultivated hydroponically<sup>4,5</sup> or on a solid medium containing the hydroponic solution (HD plates) in plastic Petri dishes<sup>28</sup> (Methods).

**Plant transformations.** Constructs were generated by cloning of PCR products into GATEWAY-compatible binary vectors (Methods; Supplementary Methods for primer sequences). *A. thaliana* was transformed by floral dip<sup>29</sup>. *Agrobacterium tumefaciens*-mediated stable transformation of *A. halleri* was performed using a tissue-culture based procedure<sup>30</sup> (Supplementary Methods). Genomic DNA and RNA gel blot analyses were performed for the primary characterization of *A. halleri* *HMA4* RNAi transformants (Supplementary Methods).

**Physiological characterization of plants.** For the determination of zinc accumulation and transcript levels, plants were cultivated hydroponically at 5  $\mu$ M ZnSO<sub>4</sub> in a controlled growth chamber for 3 weeks (Supplementary Methods). The Zn indicator Zinpyr-1 (Sigma) was used for confocal fluorescence imaging of Zn in roots of *A. halleri* and *A. thaliana* plants<sup>16</sup>. For the determination of metal tolerance, sequential root elongation assays and growth assays were performed in metal-supplemented hydroponic solutions and on HD plates, respectively (Methods).

**BAC analysis and sequencing.** Two *A. halleri* BACs (7C17, 17L07) covering the region orthologous to the *HMA4* region of the *A. thaliana* genome and containing a total of three genomic *HMA4* copies of *A. halleri*, were isolated and completely sequenced (Methods). Gene and transposable element annotations were



**Figure 4 | Characterization of *A. thaliana* expressing *AhHMA4*.** **a**, Zn localization in roots of 10-day-old seedlings. Confocal images show fluorescent Zn signals (green) and cell walls (red). Scale bars, 25  $\mu$ m; pc, pericycle; arrowheads highlight fluorescent Zn signals in xylem vessels. **b**, Relative transcript levels of *ZIP4* and *IRT3* in roots of hydroponically cultivated 5.5-week-old plants (mean  $\pm$  s.e.m.,  $n = 6$ ). **c**, Zn and Cd toxicity in shoots of 17-day-old *AhHMA4* transformants seven days after transfer to control medium (5  $\mu$ M Zn, no Cd) or medium supplemented with 150  $\mu$ M Zn or 40  $\mu$ M Cd.



made on the basis of similarity searches performed using the BLASTN and BLASTX programs against the TAIR database (<http://www.arabidopsis.org>) and the non-redundant database (nr, <http://www.ncbi.nlm.nih.gov/BLAST/>).

**Expression analyses.** Relative transcript levels were determined by real-time RT-PCR<sup>4</sup>. Histochemical GUS staining, fluorimetric quantitative GUS activity assays and *in situ* hybridizations were carried out according to standard procedures (Supplementary Methods).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 20 December 2007; accepted 28 February 2008.

Published online 20 April 2008.

- Mitchell-Olds, T. *Arabidopsis thaliana* and its wild relatives. *Trends Ecol. Evol.* **16**, 693–700 (2001).
- Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172 (2004).
- Baker, A. J. M., McGrath, S. P., Reeves, R. D. & Smith, J. A. C. in *Phytoremediation of Contaminated Soil and Water* (eds Terry, N. & Bañuelos, G. S.) 85–107 (CRC Press LLC, Boca Raton, Florida, 1999).
- Talke, I. N., Hanikenne, M. & Krämer, U. Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiol.* **142**, 148–167 (2006).
- Becher, M., Talke, I. N., Krall, L. & Krämer, U. Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *Plant J.* **37**, 251–268 (2004).
- Weber, M. et al. Comparative microarray analysis of *Arabidopsis thaliana* and *Arabidopsis halleri* roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. *Plant J.* **37**, 269–281 (2004).
- Dräger, D. B. et al. Two genes encoding *Arabidopsis halleri* MTP1 metal transport proteins co-segregate with zinc tolerance and account for high MTP1 transcript levels. *Plant J.* **39**, 425–439 (2004).
- Courbot, M. et al. A major QTL for Cd tolerance in *Arabidopsis halleri* co-localizes with HMA4, a gene encoding a heavy metal ATPase. *Plant Physiol.* **144**, 1052–1065 (2007).
- Willems, G. et al. The genetic basis of zinc tolerance in the metallophyte *Arabidopsis halleri* ssp. *halleri* (Brassicaceae): An analysis of quantitative trait loci. *Genetics* **176**, 659–674 (2007).
- Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution Int. J. Org. Evolution* **61**, 995–1016 (2007).
- Axelsen, K. B. & Palmgren, M. G. Inventory of the superfamily of P-type ion pumps in *Arabidopsis*. *Plant Physiol.* **126**, 696–706 (2001).
- Smith, N. A. et al. Total silencing by intron-spliced hairpin RNAs. *Nature* **407**, 319–320 (2000).
- Krämer, U. et al. Free histidine as a metal chelator in plants that hyperaccumulate nickel. *Nature* **379**, 635–638 (1996).
- Lasat, M. M. et al. Molecular physiology of zinc transport in the Zn hyperaccumulator *Thlaspi caerulescens*. *J. Exp. Bot.* **51**, 71–79 (2000).
- Hussain, D. et al. P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in *Arabidopsis*. *Plant Cell* **16**, 1327–1339 (2004).
- Sinclair, S. A. et al. The use of the zinc-fluorophore, Zinpyr-1, in the study of zinc homeostasis in *Arabidopsis* roots. *New Phytol.* **174**, 39–45 (2007).
- Clemens, S., Palmgren, M. G. & Krämer, U. A long way ahead: understanding and engineering plant metal accumulation. *Trends Plant Sci.* **7**, 309–315 (2002).
- Grotz, N. et al. Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proc. Natl Acad. Sci. USA* **95**, 7220–7224 (1998).
- Bert, V. et al. Genetic basis of Cd tolerance and hyperaccumulation in *Arabidopsis halleri*. *Plant Soil* **249**, 9–18 (2003).
- MacNair, M. R. et al. Zinc tolerance and hyperaccumulation are genetically independent characters. *Proc. R. Soc. Lond. B* **266**, 2175–2179 (1999).
- Windsor, A. J. et al. Partial shotgun sequencing of the *Boechera stricta* genome reveals extensive microsynteny and promoter conservation with *Arabidopsis*. *Plant Physiol.* **140**, 1169–1182 (2006).
- Küpper, H., Lombi, E., Zhao, F. J. & McGrath, S. P. Cellular compartmentation of cadmium and zinc in relation to other elements in the hyperaccumulator *Arabidopsis halleri*. *Planta* **212**, 75–84 (2000).
- Weigel, D. & Nordborg, M. Natural variation in *Arabidopsis*. How do we find the causal genes? *Plant Physiol.* **138**, 567–568 (2005).
- Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genet.* **38**, 594–597 (2006).
- Beckmann, J. S., Estivill, X. & Antonarakis, S. E. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Rev. Genet.* **8**, 639–646 (2007).
- Sugino, R. P. & Innan, H. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. *Trends Genet.* **22**, 642–644 (2006).
- Zhong, S., Khodursky, A., Dykhuizen, D. E. & Dean, A. M. Evolutionary genomics of ecological specialization. *Proc. Natl Acad. Sci. USA* **101**, 11719–11724 (2004).
- Arrivault, S., Senger, T. & Krämer, U. The *Arabidopsis* metal tolerance protein AtMTP3 maintains metal homeostasis by mediating Zn exclusion from the shoot under Fe deficiency and Zn oversupply. *Plant J.* **46**, 861–879 (2006).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1999).
- Chateau, S., Sangwan, R. S. & Sangwan-Norreel, B. S. Competence of *Arabidopsis thaliana* genotypes and mutants for *Agrobacterium tumefaciens*-mediated gene transfer: role of phytohormones. *J. Exp. Bot.* **51**, 1961–1968 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Baurain, D. Walther, C. Galante, T. Werner and the gardeners of the Max Planck Institute of Molecular Plant Physiology for assistance, R. Schmidt for *A. thaliana* 35S<sub>GUS</sub> lines, I. Somssich for pJAWOHL8, and S. Thomine for comments on the manuscript. This work was funded by: German Research Foundation Kr1967/3-1, Heisenberg Fellowship Kr1967/4-1; German Federal Ministry of Education and Research Biofuture 0311877 and GABI-ADVANCIS 0315037A; European Union RTN "METALHOME" HPRN-CT-2002-00243, InP "PHIME" FOOD-CT-2006-016253 (U.K.). Further funding was from 'Fonds spéciaux pour la Recherche, University of Liège, Belgium' (M.H., P.M.), 'Fonds de la Recherche Scientifique – FNRS', Belgium (M.H.), and the Max Planck Society (D.W.).

**Author Contributions** I.N.T., M.H., M.J.H., A.N., U.K., P.M. and J.K. performed experiments, C.L. the BAC sequencing and assembly, M.H. assembly and BAC annotation; D.W. and J.K. provided the BAC library and filters; U.K., M.H. and I.N.T. jointly designed experiments; D.W. gave experimental advice and edited the manuscript; U.K. conceived of the study and directed the research; U.K., M.H. and I.N.T. wrote and edited the manuscript; all authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). BAC sequences are available online (Genbank accession numbers EU382072, EU382073). Correspondence and requests for materials should be addressed to U.K. ([ute.kraemer@bioquant.uni-heidelberg.de](mailto:ute.kraemer@bioquant.uni-heidelberg.de)).



## METHODS

**Plant cultivation.** Plants were cultivated in a climate-controlled glasshouse at 21 °C day/17 °C night, constant humidity of 50%, with supplementary lighting from equal amounts of HPIT and Son-T Agro lamps (Philips) providing a photoperiod of 16 h, or in a growth chamber at 20 °C day/18 °C night, a humidity of 60% day/75% night, and a photoperiod of 11 h at a photon flux density  $145 \mu\text{mol m}^{-2} \text{s}^{-1}$ .

**Generation of constructs.** All *HMA4* sequences were PCR-amplified using a proofreading polymerase (Pfu Turbo, Stratagene), and verified by sequencing after directional cloning into pENTR/D TOPO (Invitrogen) and again after site-directed recombination into the respective GATEWAY-compatible binary vector before plant transformation, unless indicated otherwise.

To generate the *HMA4* RNAi construct for transformation of *A. halleri*, a fragment corresponding to bp 2,541–2,997 of the open reading frame of *AhHMA4* was amplified from *A. halleri* cDNA and subcloned into the binary vector pJAWOHL8 (GenBank accession AF408413), generating an intron-spliced hairpin construct with antisense *AhHMA4* fragment – intron – sense *AhHMA4* fragment configuration downstream of the CaMV 35S promoter<sup>12</sup>.

To generate promoter-reporter constructs for the transformation of *A. thaliana* and *A. halleri*, an *AtHMA4* promoter fragment of 2,625 bp was amplified from genomic DNA of *A. thaliana*. The *AhHMA4-1*, *AhHMA4-2* and *AhHMA4-3* promoter fragments of 2,326, 1,311 and 1,082 bp, respectively, were amplified from *A. halleri* genomic DNA. All promoter fragments, which include promoter and 5'-UTR regions and end after the first 30 bp of the respective *HMA4* coding sequence, were subcloned into the pMDC163 binary vector<sup>31</sup>.

The *AhHMA4<sub>p</sub>-HMA4* construct for transformation of *A. thaliana* was generated by inserting both the *AhHMA4-1* promoter fragment and the full-length *AhHMA4* coding sequence into a promoter-less variant of the pMDC32 vector<sup>31</sup> (Supplementary Methods).

**Physiological characterization of plants.** *A. halleri* plants were maintained in hydroponic culture<sup>4,5</sup> in a climate-controlled glasshouse. *A. halleri* *HMA4* RNAi lines and wild type control individuals were propagated vegetatively via rooting of shoot cuttings on sand for 5 weeks and subsequently cultivated in hydroponics for 4–5 weeks before initiating experimental treatments. Results for wild type controls were averaged from the following *A. halleri* individuals: (1) untransformed, (2) regenerated from tissue culture following mock transformation and (3) transformed with a cauliflower mosaic virus (CaMV) 35S promoter-*GUS*-*Intron* construct<sup>32</sup>. *A. thaliana* plants were grown from seeds in hydroponic culture<sup>4,5</sup> for 6 weeks before experimental treatments were initiated.

Metal tolerance of *A. halleri* genotypes was determined in a climate-controlled glasshouse by measuring root elongation in a slightly modified hydroponic control solution (5  $\mu\text{M}$  Zn, no added Cd, 0.14 mM  $\text{KH}_2\text{PO}_4$  to avoid precipitation at high metal concentrations), and subsequently in the presence of sequentially increased metal concentrations, in 5-day intervals (Zn: 800  $\mu\text{M}$ , 1.5 mM; Cd: 10  $\mu\text{M}$ , 30  $\mu\text{M}$ )<sup>20</sup>. Two experiments were performed for each metal.

For metal tolerance assays in *A. thaliana* *AhHMA4* transformants (T1 generation), 10-day-old hygromycin-resistant seedlings were transferred onto HD plates containing 5  $\mu\text{M}$   $\text{ZnSO}_4$  (control), 150  $\mu\text{M}$   $\text{ZnSO}_4$ , or 40  $\mu\text{M}$   $\text{CdSO}_4$  (and 5  $\mu\text{M}$   $\text{ZnSO}_4$ ) and maintained in a climate-controlled growth chamber. Photographs of representative individuals were taken one week after transfer. The experiment was repeated three times, with three plates per genotype and 10 replicate seedlings per plate in each experiment.

For Zn imaging, segments of newly formed distal roots were used of *A. halleri* plants grown in a hydroponic medium containing 1  $\mu\text{M}$   $\text{ZnSO}_4$ . For *A. thaliana*, seedlings (T3 generation) were grown for 10 days on HD plates in a

climate-controlled growth chamber and incubated in hydroponic solution with 25  $\mu\text{M}$   $\text{ZnSO}_4$  for 2.5 h before staining. Fluorescence of the Zn-Zinpyr-1 chelate was observed using confocal laser scanning microscopy (Leica SP2) with 488 nm excitation using FITC and Texas Red filters. For *A. thaliana*, two transgenic *AhHMA4* lines showing increased *HMA4* transcript levels were analysed. For *A. halleri*, *AhHMA4* RNAi lines 4.2.1, 5.4.2 and 6.5.1 were analysed (4.2.1 is shown in Fig. 1d). Images shown are representative of the results from at least 6 wild type plants and 6 transgenic plants.

**Real-time RT-PCR analysis.** Preparation of total DNase-treated RNA, synthesis of cDNA, primer design and sequences, quality control and data analysis were as described<sup>4</sup>. Values shown in Fig. 1a are averages from three to four independent biological experiments, values shown in Figs 1e, f, 2b and 4b are averages from six to eight technical replicates from one experiment representative of a total of three independent biological experiments.

**BAC analysis and sequencing.** A BAC library was constructed in the pIndigoBAC-536 vector using *Hind*III-digested total genomic DNA of 8 individuals from *A. halleri* ssp. *halleri*, accessions Roderbacherbrunn and Stutenkamm (Germany) (Keygene), and screened with an *HMA4* probe<sup>4</sup> as described<sup>33</sup>. A total of 18 positive BAC clones were obtained. A combination of Southern blots and restriction pattern analyses, together with BAC end sequencing, demonstrated that two BACs (7C17, 17L07) covered all of a total of three genomic *HMA4* copies of *A. halleri*, and the region orthologous to the *HMA4* region of the *A. thaliana* genome.

For complete sequencing, BAC 7C17 and 17L07 subclone libraries were prepared with an average insert size of 5–10 kb, and a total of 1,152 positive clones for each BAC were end-sequenced to a 13-fold coverage as described<sup>34</sup>. Sequence reads were base-called using the LifeTrace software<sup>35</sup> and assembled into contigs with the PHRAP and CONSED software packages (<http://www.phrap.org>)<sup>36</sup>. Remaining gaps and ambiguously assembled regions were closed and polished by primer walking and PCR with oligonucleotides designed based on the flanking sequences.

Comparative genomic sequence analyses were done with the GenomeVISTA software (<http://genome.lbl.gov/vista/index.shtml>)<sup>37,38</sup> to identify syntenic regions.

**Statistics.** All data evaluation and statistics were done using Microsoft Excel. Statistical analysis of data from *A. halleri* *HMA4* RNAi lines was performed by multiple comparisons based on ANOVA.

31. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes in plants. *Plant Physiol.* **133**, 462–469 (2003).
32. Vancanneyt, G. et al. Construction of an intron-containing marker gene: splicing of the intron in transgenic plants and its use in monitoring early events in *Agrobacterium*-mediated plant transformation. *Mol. Gen. Genet.* **220**, 245–250 (1990).
33. Benderoth, M. et al. Positive selection driving diversification in plant secondary metabolism. *Proc. Natl Acad. Sci. USA* **103**, 9118–9123 (2006).
34. Eppinger, M. et al. Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet.* **2**, e120 (2006).
35. Walther, D., Bartha, G. & Morris, M. Basecalling with LifeTrace. *Genome Res.* **11**, 875–888 (2001).
36. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
37. Bray, N., Dubchak, I. & Pachter, L. AVID: A global alignment program. *Genome Res.* **13**, 97–102 (2003).
38. Couronne, O. et al. Strategies and tools for whole-genome alignments. *Genome Res.* **13**, 73–80 (2003).



## LETTERS

## Human metabolic phenotype diversity and its association with diet and blood pressure

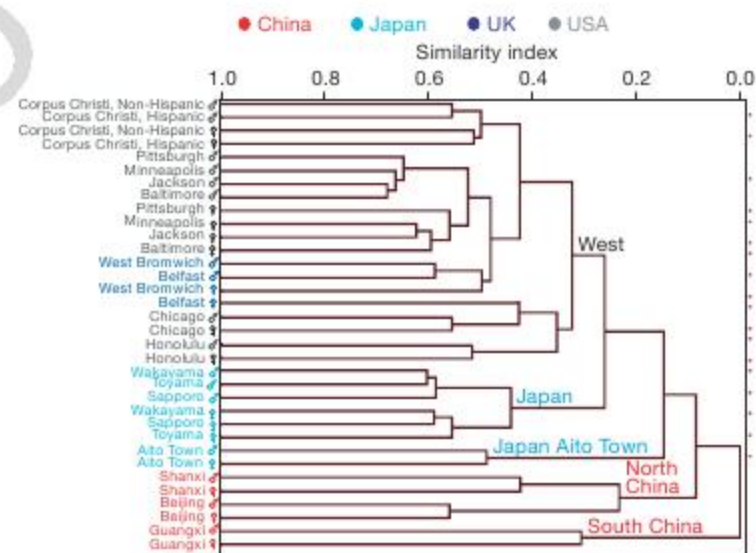
Elaine Holmes<sup>1\*</sup>, Ruey Leng Loo<sup>1,2\*</sup>, Jeremiah Stamler<sup>3</sup>, Magda Bictash<sup>1,2</sup>, Ivan K. S. Yap<sup>1,2</sup>, Queenie Chan<sup>2</sup>, Tim Ebbels<sup>1</sup>, Maria De Iorio<sup>2</sup>, Ian J. Brown<sup>2</sup>, Kirill A. Veselkov<sup>1</sup>, Martha L. Daviglus<sup>3</sup>, Hugo Kesteloot<sup>4</sup>, Hirotsugu Ueshima<sup>5</sup>, Liancheng Zhao<sup>6</sup>, Jeremy K. Nicholson<sup>1</sup> & Paul Elliott<sup>2</sup>

Metabolic phenotypes are the products of interactions among a variety of factors—dietary, other lifestyle/environmental, gut microbial and genetic<sup>1–3</sup>. We use a large-scale exploratory analytical approach to investigate metabolic phenotype variation across and within four human populations, based on <sup>1</sup>H NMR spectroscopy. Metabolites discriminating across populations are then linked to data for individuals on blood pressure, a major risk factor for coronary heart disease and stroke (leading causes of mortality worldwide<sup>4</sup>). We analyse spectra from two 24-hour urine specimens for each of 4,630 participants from the INTERMAP epidemiological study<sup>5</sup>, involving 17 population samples aged 40–59 in China, Japan, UK and USA. We show that urinary metabolite excretion patterns for East Asian and western population samples, with contrasting diets, diet-related major risk factors, and coronary heart disease/stroke rates, are significantly differentiated ( $P < 10^{-16}$ ), as are Chinese/Japanese metabolic phenotypes, and subgroups with differences in dietary vegetable/animal protein and blood pressure<sup>6</sup>. Among discriminatory metabolites, we quantify four and show association ( $P < 0.05$  to  $P < 0.0001$ ) of mean 24-hour urinary formate excretion with blood pressure in multiple regression analyses for individuals. Mean 24-hour urinary excretion of alanine (direct) and hippurate (inverse), reflecting diet and gut microbial activities<sup>2,7</sup>, are also associated with blood pressure of individuals. Metabolic phenotyping applied to high-quality epidemiological data offers the potential to develop an area of aetiopathogenetic knowledge involving discovery of novel biomarkers related to cardiovascular disease risk.

Prehypertensive and hypertensive blood pressure (BP) is prevalent among a majority of middle-aged and older adults in most countries, and is a major risk factor perpetuating the cardiovascular disease epidemic<sup>8</sup>. The goal of the INTERMAP metabonomic study<sup>1</sup> is to develop metabolic phenotyping approaches to elucidate aetiopathogenetic mechanisms underlying the global BP problem and related disorders<sup>9</sup>. It aims to identify urinary metabolites that discriminate across population/subgroup strata defined by geographic or dietary criteria, and assess—for individuals—relationships of these metabolites to BP. The basic concepts are: (1) for population/subgroup strata with differing coronary heart disease (CHD)/stroke rates or BP levels, the differences are largely attributable to lifestyles, especially diet; (2) these differences across strata are reflected in urinary metabolite patterns and specific metabolic biomarkers; and (3) for individuals, these biomarkers may relate independently to their BP. We use a technology platform that is analytically unbiased and

detects a wide variety of metabolites from dietary, gut microbial and host metabolism sources in one analytical sweep<sup>1</sup>, thus maximizing opportunity for novel biomarker discovery. The numbers of individuals in our population samples are as follows: China,  $n = 832$ ; Japan,  $n = 1,138$ ; UK,  $n = 496$ ; and USA,  $n = 2,164$ .

A schematic summarizing data-analysis strategy is shown in Supplementary Fig. 1a, b. From hierarchical clustering analysis (HCA) with group average linkage applied to the probabilistic quotient normalized<sup>9</sup> median <sup>1</sup>H NMR spectrum (Methods), we find that East Asian and western populations have well-differentiated metabolic phenotypes (Fig. 1). Results for first and second urine specimens show highly similar clustering order (Fig. 1 and Supplementary Fig. 2a), as do HCA dendrograms generated using the single-linkage method (Supplementary Fig. 2b, c). Geographic metabolic differences are greater than gender



**Figure 1 | Hierarchical cluster analysis using group average linkage based on median <sup>1</sup>H NMR urine spectra, by population sample and gender ( $n = 4,630$ ).** Data for first 24-h urinary specimens. The hierarchical cluster analysis (HCA) algorithm produces a dendrogram showing the overall similarity/dissimilarity between population samples. Similarity index is normalized to intercluster distance. The similarity index measures the multivariate distance between clusters. A similarity of one indicates zero distance between clusters; a value of zero indicates the maximum intercluster separation seen in the data. Each branch of the dendrogram defines a subcluster; population samples within subclusters are more similar to each other than to those in other subclusters.

<sup>1</sup>Biomolecular Medicine, Division of Surgery, Oncology, Reproductive Biology and Anaesthetics (SORA), Faculty of Medicine, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. <sup>2</sup>Department of Epidemiology and Public Health, Imperial College London, St Mary's Campus, London W2 1PG, UK. <sup>3</sup>Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. <sup>4</sup>Department of Public Health, Division of Epidemiology, Akademisch Ziekenhuis St Rafael, Leuven B-3000, Belgium. <sup>5</sup>Department of Health Science, Shiga University of Medical Science, Otsu, Shiga 520-2192, Japan. <sup>6</sup>Department of Epidemiology, Fu Wai Hospital and Cardiovascular Institute, Chinese Academy of Medical Sciences, Beijing 100037, China.

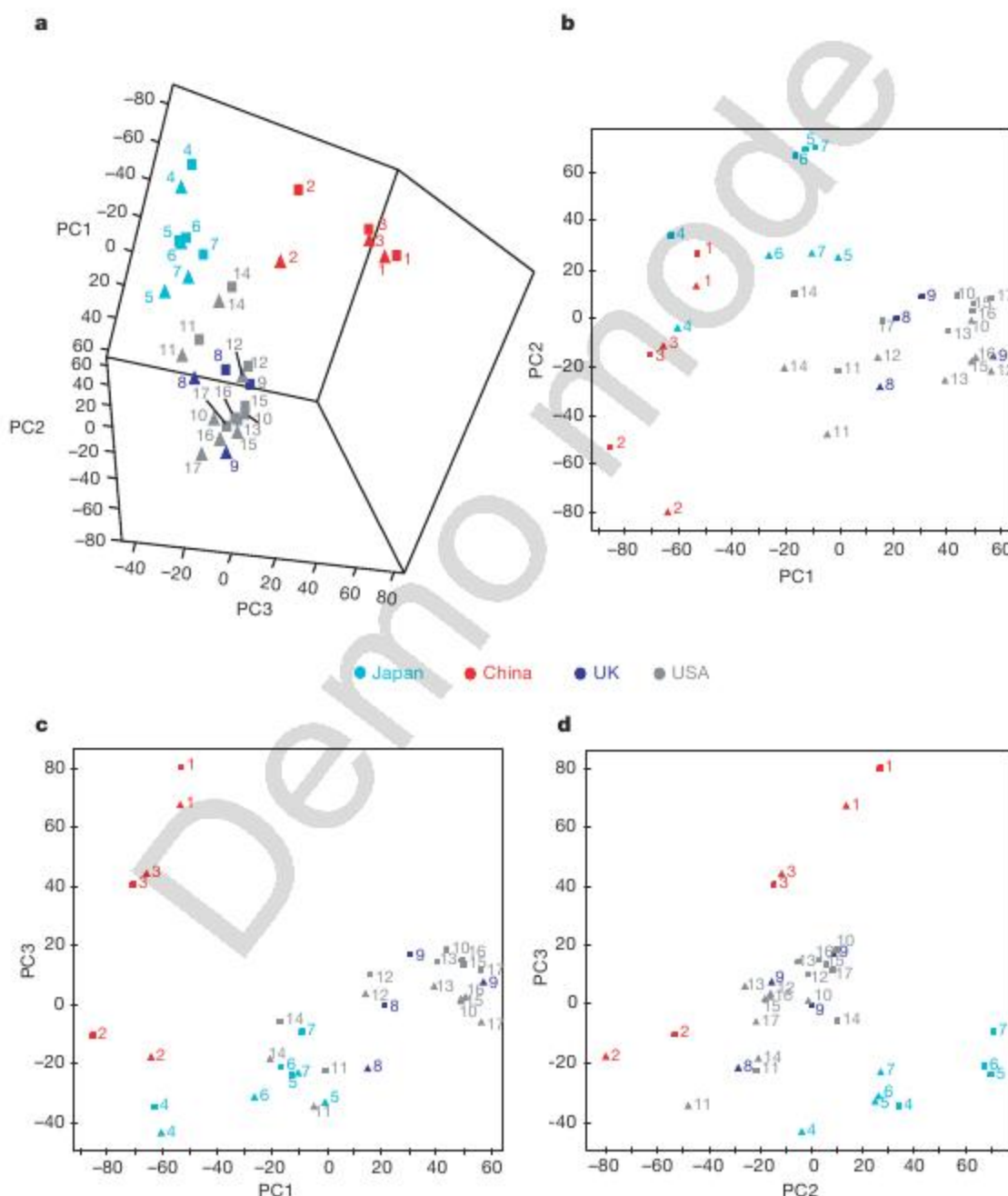
\*These authors contributed equally to this work.



differences. Metabolic phenotypes of southern (Guangxi) and northern (Beijing and Shanxi) Chinese are also differentiated; those of UK and USA population samples overlap. These findings are consistent with principal components analysis (PCA) results (Fig. 2a–d and Supplementary Fig. 3a–d). The plots of PCA scores show similarity of median urine metabolite profiles by gender, with separate subclusters for China (north and south), Japan and the two western population samples. Analyses of spectroscopic data sets from first and second urine specimens are highly consistent (Supplementary Fig. 4a), as are analyses limited to normal weight, non-diabetic participants (Supplementary Fig. 4b), and analyses repeated after removal of metabolic outliers (individual data shown in Supplementary Fig. 5) using the 95% Hotelling's  $T^2$  statistic<sup>10</sup> (Methods and Supplementary Fig. 6a, b).

We examine pairwise comparisons across countries (except for UK and USA which are poorly discriminated) and show significant

differences (Hotelling's  $T^2$ ,  $P < 10^{-16}$ ) using orthogonally filtered partial least squares discriminant analysis (O-PLS-DA)<sup>11</sup>, after exclusion of metabolic outliers. Japanese living in Japan and Japanese-Americans are also differentiated ( $P < 10^{-16}$ ). Discriminatory metabolites (Supplementary Table 1a, b) are identified from the O-PLS-DA coefficients on the basis of four criteria:  $P$  value, the rank of the  $P$  value, stability of rank and regression coefficient strength from a bootstrap resampling procedure (Methods). They include metabolites of predominantly dietary origin, for example, amino acids, creatine and trimethylamine- $N$ -oxide; compounds related to energy metabolism (acetylcarnitine, tricarboxylic acid cycle intermediates); and dicarboxylic acids (for example, suberate). We also find that population gut microbial-mammalian co-metabolites<sup>2</sup> are discriminatory, for example, hippurate, phenylacetylglutamine and methylamines; we have previously shown structural differences in



Key: 1, Beijing; 2, Guangxi; 3, Shanxi; 4, Aito Town; 5, Sapporo; 6, Toyama; 7, Wakayama; 8, Belfast; 9, West Bromwich; 10, Baltimore; 11, Chicago; 12, Corpus Christi Hispanic; 13, Corpus Christi non-Hispanic; 14, Honolulu; 15, Jackson; 16, Minneapolis; 17, Pittsburgh.

**Figure 2 | Plots of cross-validated principal components analysis scores ( $n = 4,630$ ).** a, Pseudo three-dimensional plot for principal components (PC) 1–3; b, PC2 versus PC1; c, PC3 versus PC1; d, PC3 versus PC2. Median  $^1\text{H}$  NMR spectra of the first 24-h urine specimens stratified by country and by gender, female (triangles) and male (squares).  $R^2_x = 74.2\%$  (percentage

variation in the NMR data explained by the model);  $Q^2_x = 49.6\%$  (percentage variation in the NMR data predictable by the model from cross-validation). The cross-validated scores values for the first three components are available in Supplementary Information. Symbols in b–d as in a.



Chinese and American gut microbial speciation and direct linkage of microbial composition to metabolic phenotype<sup>12</sup>.

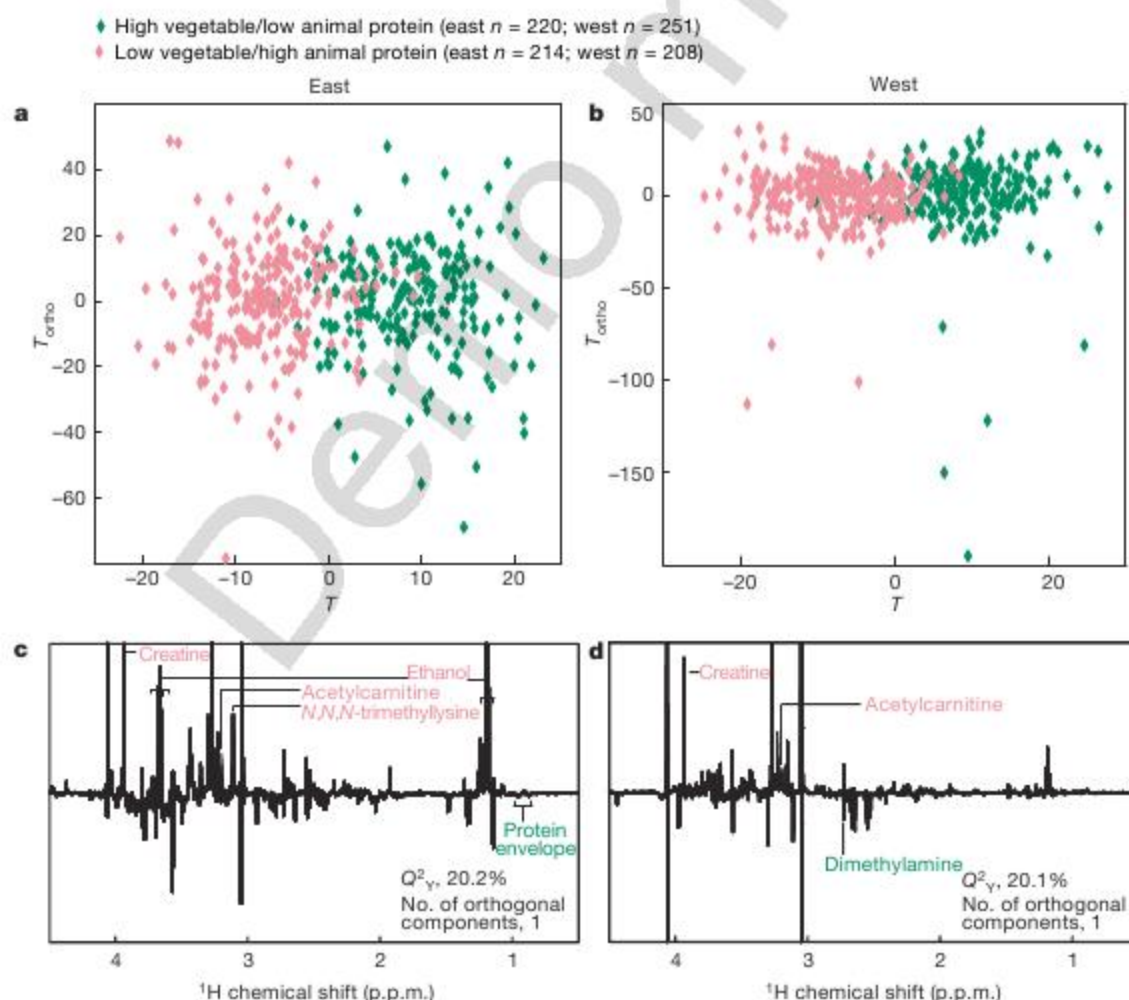
Participants consuming dietary protein predominantly from vegetable or from animal sources (subgroups differing in BP levels<sup>6</sup>) are also differentiated (Hotelling's  $T^2$   $P < 10^{-13}$ ) for East Asian and western samples considered separately (Fig. 3a, b and Supplementary Fig. 7a, b). Significant discriminatory metabolites (Fig. 3c, d and Supplementary Fig. 7c, d) closely correspond to those from the pairwise country comparisons.

We then quantify four discriminatory metabolites (alanine, formate, hippurate and *N*-methylnicotinate) from the  $^1\text{H}$  NMR spectra (Methods and Supplementary Fig. 1b) and analyse these with respect to all other spectral variables using O-PLS regression<sup>11</sup>. Largest  $r^2$  values (first urine specimens), other than for intra-molecular correlations, are (1) for alanine, with 2-oxoglutarate, reflecting close metabolic linkage via glutamate-pyruvate transaminase activity<sup>13</sup>; (2) for formate, with alanine, explained by pyruvate/Co-A metabolism; (3) for hippurate with *N*-methylnicotinate; and (4) for *N*-methylnicotinate with hippurate, reflecting common or related renal transporter/secretion mechanisms. We also find significant correlations ( $|r| \geq 0.10$ ,  $P < 10^{-9}$ ) with other variables (Supplementary Table 2): (1) alanine, positively with energy intake, dietary cholesterol, body mass index, 24-h urinary  $\text{Na}^+$  and  $\text{K}^+$  and  $\text{Na}^+/\text{K}^+$  ratio; inversely with alcohol intake; (2) formate, positively with energy intake, 24-h urinary  $\text{Na}^+$  and  $\text{K}^+$ ; (3) hippurate, positively with dietary fibre,  $\text{Mg}^{2+}$ , phosphorus, 24-h urinary  $\text{Na}^+$  and  $\text{K}^+$ ; inversely with alcohol intake and urinary  $\text{Na}^+/\text{K}^+$  ratio; and (4) *N*-methylnicotinate, positively with dietary  $\text{Mg}^{2+}$ , 24-h urinary  $\text{Na}^+$

and  $\text{K}^+$ ; inversely with urinary  $\text{Na}^+/\text{K}^+$  ratio. Strongest correlations are with 24-h urinary  $\text{Na}^+$  excretion for alanine (Pearson  $r = 0.39$ ) and formate ( $r = 0.37$ ), and with 24-h urinary  $\text{K}^+$  excretion for hippurate ( $r = 0.40$ ) (when excreted as the sodium salt, hippurate can cause an aldosterone-mediated increase in  $\text{K}^+$  excretion<sup>14</sup>).

In multiple linear regression models (four per metabolite for each of systolic and diastolic BP), accounting for the key non-dietary and dietary/urinary excretion variables associated with BP<sup>15</sup>, we find significant inverse associations of formate with both systolic and diastolic BP (all eight models, Table 1); also of hippurate in six models, and a significant direct association of alanine with BP in five models (Table 1). Regression estimates are similar or larger with analyses restricted to 'non-intervened' individuals<sup>16</sup>, that is, people without special diet/nutritional supplements or diagnosis/treatment for cardiovascular disease or diabetes (Supplementary Table 3). Technical errors and reliability estimates of quantified metabolites are provided in Supplementary Information and Supplementary Table 4.

Using large-scale metabolic phenotyping, we have identified novel candidate urinary biomarkers related to BP. Endogenous formate is largely the product of one-carbon metabolism via the activities of mitochondrial and cytosolic serine hydroxymethyl transferases, and the tetrahydrofolate pathway<sup>17</sup>. Formate is also produced as one by-product of fermentation of dietary fibre by the gut microbiome<sup>18</sup>. It is involved in active  $\text{Cl}^-$  reabsorption at the apical proximal tubule via the CFEX anion exchanger under inhibitory control of the serine-threonine kinase WNK4; gain-of-function mutations in WNK4 cause pseudohypoaldosteronism type II, a mendelian disorder associated



**Figure 3 | O-PLS-DA scores and loadings plots (bootstrap analyses) for participants reporting high vegetable/low animal protein and low vegetable/high animal protein intakes, first 24-h urinary specimens.** Plots (one orthogonal component) compare top and bottom quartiles, adjusted for sample, age and sex, from a, East Asian, and b, western population samples. Loadings plots from the O-PLS-DA bootstrap analyses are shown with discriminatory metabolites labelled (see Methods for metabolite

selection criteria) for c, East Asian and d, western participants. Analyses are after removal of metabolic outliers using the 95% Hotelling's  $T^2$  statistic in the initial PCA. The plots show the number of participants, the number of components used in each model and the  $Q^2$  values (percentage variation in the protein subgroup assignment predictable by the model from cross-validation).



**Table 1 | Estimated mean differences in systolic and diastolic BP**

Urinary metabolite	A*				B†			
	Not adjusted for BMI‡		Adjusted for BMI‡		Not adjusted for BMI‡		Adjusted for BMI‡	
	Systolic blood pressure (mm Hg)							
Alanine	2.69	(6.06)	0.40	(0.92)	2.66	(5.54)	1.13	(2.43)
Formate	-1.19	(-2.62)	-1.42	(-3.29)	-1.94	(-3.92)	-1.04	(-2.20)
Hippurate	-2.10	(-4.85)	-1.63	(-3.95)	-1.72	(-3.70)	-0.82	(-1.83)
N-methylnicotinate	-0.09	(-0.21)	0.20	(0.49)	0.00	(0.00)	0.65	(1.53)
	Diastolic blood pressure (mm Hg)							
Alanine	1.57	(5.17)	0.17	(0.55)	1.58	(4.77)	0.61	(1.90)
Formate	-0.90	(-2.96)	-1.02	(-3.49)	-1.41	(-4.22)	-0.86	(-2.65)
Hippurate	-0.98	(-3.33)	-0.71	(-2.50)	-0.77	(-2.42)	-0.23	(-0.73)
N-methylnicotinate	-0.07	(-0.25)	0.09	(0.32)	-0.01	(-0.03)	0.37	(1.27)

Systolic and diastolic blood pressure differences per +2 s.d. difference in each of four quantified urinary metabolites (mean of two 24-h urine values). Numbers in parentheses are Z scores, that is, regression coefficient divided by standard error (Z-score  $\geq 1.96$ ,  $P < 0.05$ ;  $\geq 2.58$ ,  $P < 0.01$ ;  $\geq 3.29$ ,  $P < 0.001$ ;  $\geq 3.89$ ,  $P < 0.0001$ ). 2 s.d. difference for alanine = 0.34 mmol per 24 h ( $n = 4,232$ ); formate = 0.29 mmol per 24 h ( $n = 4,147$ ); hippurate = 3.55 mmol per 24 h ( $n = 4,184$ ); N-methylnicotinate = 0.41 mmol per 24 h ( $n = 4,081$ ) (chemical shifts used for quantification: alanine,  $\delta$  1.48; formate,  $\delta$  8.45; hippurate,  $\delta$  7.85 and N-methylnicotinate,  $\delta$  4.44). Regression coefficients for individuals are pooled across countries (Methods); there is no evidence for cross-country heterogeneity in size of coefficients.

\*A: Adjusted for age, sex, sample, special diet, supplement use, cardiovascular disease or diabetes mellitus diagnosis, physical activity (h per 24 h moderate or heavy activity), family history of high blood pressure.

†B: A + 7-day alcohol (g per 24 h) + urinary Na<sup>+</sup> (mmol per 24 h) + urinary K<sup>+</sup> excretion (mmol per 24 h).

‡Body mass index (kg m<sup>-2</sup>).

with hypertension<sup>19</sup>. We show that urinary formate and urinary Na<sup>+</sup> excretion are positively correlated. Given the central importance of NaCl in control of BP and the rise of BP with age<sup>15,20</sup>, our findings suggest a previously unrecognized role for formate in BP regulation.

The inverse association of hippurate (benzoyl glycine) with BP may reflect physiological connections with diet<sup>7</sup> and gut microbial activity<sup>2</sup>. Availability of calories from the diet is also modulated by gut microbes in human obesity<sup>21</sup>, which in turn relates to BP<sup>15</sup>. We previously reported that dietary alanine is higher in people consuming a predominantly animal compared with a predominantly vegetable diet<sup>6</sup>, consistent with our findings here of a direct association of urinary alanine excretion with BP. Also in experimental animal models, alanine modulates cardiovascular responses to circulating catecholamines and increases BP<sup>22</sup>.

Cross-population metabolic differences shown here add a new dimension to the decades-long knowledge of East–West contrasting patterns of diet, diet-related major risk factors and CHD/stroke mortality (Supplementary Information, Supplementary Figs 8a, b and 9a, b, and Supplementary Tables 5 and 6). We have shown that urinary metabolic phenotyping across populations/subgroups at differing risks of CHD/stroke and high BP identifies novel candidate biomarkers that relate to BP of individuals. This may provide the basis for a new ‘metabolome-wide association’ approach in molecular epidemiology to help understand the complex interactions of lifestyles, environment and genes that determine major diseases in the twenty-first century.

## METHODS SUMMARY

INTERMAP is an international standardized population-based epidemiological investigation of diet and BP<sup>5</sup>. We collected four in-depth 24-h multi-pass dietary recalls, eight BP measurements, anthropometric and questionnaire data and obtained two timed 24-h urine specimens, on average three weeks apart, from each individual according to standard protocol. We performed 600 MHz <sup>1</sup>H NMR spectroscopy on these samples; estimated within-specimen reproducibility was >98% from blinded analysis of 8% specimens split in the field<sup>1</sup>. Scans (64) for each spectrum were acquired using standard parameters and pre-processing algorithms<sup>23</sup>; spectra were reduced to 7,100 variables by integrating spectral intensity in segments (width in chemical shift  $\delta$ , 0.001) corresponding to the regions  $\delta = 0.5$ – $9.5$  (excluding  $\delta = 4.5$ – $6.4$  containing the residual water and urea resonances). We performed HCA and PCA using the median NMR spectrum for each of 34 gender-specific population samples. The Hotelling  $T^2$  statistic (95% criterion)<sup>10</sup>, calculated from PCA analyses of all first and second 24-h urine spectra, was used to remove metabolic outliers ( $n = 575$ ), enabling finer spectral detail. For the remaining 4,055 individuals, we used O-PLS-DA<sup>11</sup> to detect patterns of metabolites differentiating pairs of populations/subgroups. All models were computed separately for first and second 24-h urine specimens. For four quantified metabolites, we estimated technical error<sup>5</sup> from the 8% split samples and intra-individual reliability<sup>24</sup> from comparison of first and second

24-h urinary excretion values. Correlation-regression analyses were performed using standard methods<sup>6</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 December 2007; accepted 3 March 2008.

Published online 20 April 2008.

- Dumas, M. E. et al. Assessment of analytical reproducibility of <sup>1</sup>H NMR spectroscopy based metabolomics for large-scale epidemiological research: the INTERMAP study. *Anal. Chem.* **78**, 2199–2208 (2006).
- Nicholson, J. K., Holmes, E. & Wilson, I. D. Gut microorganisms, mammalian metabolism and personalized health care. *Nature Rev. Microbiol.* **3**, 431–438 (2005).
- Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
- Murray, C. J. & Lopez, A. D. Mortality by cause for eight regions of the world: global burden of disease study. *Lancet* **349**, 1269–1276 (1997).
- Stamler, J. et al. INTERMAP: Background, aims, design, methods, and descriptive statistics (non-dietary). *J. Hum. Hypertens.* **17**, 591–608 (2003).
- Elliott, P. et al. Association between protein intake and blood pressure: the INTERMAP study. *Arch. Intern. Med.* **166**, 79–87 (2006).
- Mulder, T. P., Rietveld, A. G. & van Amelsvoort, J. M. Consumption of both black tea and green tea results in an increase in the excretion of hippuric acid into urine. *Am. J. Clin. Nutr.* **81** (Suppl.), 256S–260S (2005).
- Elliott, P. & Stamler, J. in *Coronary Heart Disease Epidemiology: From Aetiology to Public Health* 2nd edn (eds Marmot, M. & Elliott, P.) 751–768 (Oxford Univ. Press, Oxford, UK, 2005).
- Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabolomics. *Anal. Chem.* **78**, 4281–4290 (2006).
- Hotelling, H. The generalization of Student's ratio. *Ann. Math. Stat.* **2**, 360–378 (1931).
- Trygg, J. & Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* **16**, 119–128 (2002).
- Li, M. et al. Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl Acad. Sci. USA* **105**, 2117–2122 (2008).
- Chen, S.-H. & Giblett, E. R. Polymorphism of soluble glutamic-pyruvic transaminase: a new genetic marker in man. *Science* **173**, 148–149 (1971).
- Lin, S.-H., Lin, Y.-F. & Halperin, M. L. Hypokalaemia and paralysis. *Q. J. Med.* **94**, 133–139 (2001).
- Intersalt Co-operative Research Group. Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. *Br. Med. J.* **297**, 319–328 (1988).
- Elliott, P. et al. Dietary phosphorus and blood pressure. International study of macro- and micro-nutrients and blood pressure. *Hypertension* **51**, 669–675 (2008).
- Gregory, J. F. et al. Primed, constant infusion with 2H3 serine allows *in vivo* kinetic measurement of serine turnover, homocysteine remethylation and transsulfuration processes in human one-carbon metabolism. *Am. J. Clin. Nutr.* **72**, 1535–1541 (2000).
- Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc. Natl Acad. Sci. USA* **103**, 10011–10016 (2006).
- Kahle, K. T. et al. WNK4 regulates apical and basolateral Cl<sup>−</sup> flux in extrarenal epithelia. *Proc. Natl Acad. Sci. USA* **101**, 2064–2069 (2004).



20. Elliott, P. *et al.* Change in salt intake affects blood pressure of chimpanzees: Implications for human populations. *Circulation* **116**, 1563–1568 (2007).
21. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
22. Conlay, L. A., Maher, T. J. & Wurtman, R. J. Alanine increases blood pressure during hypotension. *Pharmacol. Toxicol.* **66**, 415–416 (1990).
23. Holmes, E. *et al.* Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. *Anal. Chem.* **79**, 2629–2640 (2007).
24. Grandits, G. A. *et al.* Method issues in dietary data analysed in the Multiple Risk Factor Intervention Trial. *Am. J. Clin. Nutr.* **65** (Suppl.), 211S–227S (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** INTERMAP is supported by the US National Heart, Lung, and Blood Institute (RO1 HL50490 and RO1 HL084228); the Chicago Health Research Foundation; and national agencies in Japan (the Ministry of Education, Science, Sports, and Culture), China and the UK. The funders had no role in the design and

conduct of the study, or in the collection, management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript. The INTERMAP study has been accomplished through the work of the staff at the local, national and international centres. A partial listing of colleagues is in ref. 5. We thank M. Rantalainen, O. Cloarec, E. Want and O. Beckonert (Imperial College London) for their assistance with the statistical and NMR analyses; and P. Oefner and H. Kaspar (University of Regensburg) for gas chromatography mass spectrometry analyses.

**Author Contributions** The INTERMAP study was conceived by J.S., P.E. and Rose Stampler (deceased); INTERMAP urinary amino acids study was by J.S., P.E., M.L.D. and H.K.; INTERMAP metabolomics study was by J.K.N. and P.E., with E.H., and J.S., M.L.D. The manuscript was written by P.E., J.K.N., E.H. and J.S.; analyses were done by R.L.L., M.B., I.K.S.Y., Q.C. and I.J.B. T.E., M.D.I., and K.V. provided statistical and analytical support. H.U. and L.Z. were responsible for data collection. All authors reviewed and approved the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.K.N. ([j.nicholson@imperial.ac.uk](mailto:j.nicholson@imperial.ac.uk)) or P.E. ([p.elliott@imperial.ac.uk](mailto:p.elliott@imperial.ac.uk)).



## METHODS

**Storage, preparation and  $^1\text{H}$  NMR spectroscopic analysis of urine.** Urine aliquots (containing boric acid preservative) were stored at  $-30^\circ\text{C}$ , thawed before use, and  $^1\text{H}$  NMR spectra acquired using a standard one-dimensional pulse sequence with water suppression (Bruker Avance 600 spectrometer operating at 600.29 MHz in flow-injection mode<sup>23</sup>). Spectra were analysed in segments of width 0.6 Hz (less than the 1 Hz frequency resolution), thus retaining all structural information. We normalized the data to total spectral area to remove outliers. In all other models, the probabilistic quotient method<sup>9</sup> was used to expose finer detail in the metabolic profile, as it is relatively unaffected by residual outlying samples.

**Data analyses.** Data were available for 4,630 of the 4,680 INTERMAP participants (Supplementary Fig. 1a). For population/subgroup analyses, we used data separately from the first and second urine specimens since averaging them loses spectral resolution due to slight shifts in peak registration. For individual-level analyses, for example, regression with BP, we used the mean of the two 24-h urinary values to increase precision<sup>24</sup>.

**HCA and PCA.** We used the median spectrum of each gender-specific population sample for each of the two 24-h urine specimens. For HCA we used Pirouette (version 3.1.1, Infometrix Inc.) with euclidean distance, and both group average and single linkages. We computed PCA models in Simca P+ (version 11, Umetrics) using sevenfold cross-validation<sup>25</sup> to select the number of components.

**Detection and removal of metabolic outliers.** PCA modelling of all 4,630 participants revealed outlying groups due to high levels of urinary glucose, trimethylamine-N-oxide, ethanol, acetaminophen and their metabolites (Supplementary Fig. 5). To remove outliers, we normalized the spectra to total area and applied Pareto scaling (dividing each variable by the square root of the standard deviation). This method weights the variables such that high concentration metabolites do not dominate the model, while avoiding noise amplification. Participants whose scores mapped outside of the 95% Hotelling  $T^2$  ellipse<sup>10</sup> in either first or second specimens were excluded ( $n = 575$ ), leaving  $n = 4,055$  individuals.

**O-PLS-DA<sup>11</sup> and selection of discriminatory metabolites.** We used an in-house O-PLS-DA algorithm in MATLAB 7.3.1 (MathWorks) to establish pairwise models between populations/subgroups, with variables scaled to unit variance. We used bootstrap resampling to identify discriminatory variables. Our method of metabolite identification used results from urinary collections: a metabolite was only considered discriminatory (Supplementary Table 1) if it was significantly associated and ranked among the top metabolites for both specimens. At each iteration, a bootstrap sample of the same size as the full sample was constructed by sampling at random with replacement. The sample was used to compute an O-PLS-DA model and the corresponding regression coefficients  $b_i$  were obtained, representing the contribution of the  $i$ th metabolic variable to between-group discrimination. This procedure was repeated 250 times; the resulting sampling distribution generated the bootstrap standard deviation of each regression coefficient,  $b_{i\text{b}}$ , for calculation of an approximate Student's  $t$  statistic and corresponding  $P$  value,  $P_i$ , for the significance of the  $i$ th coefficient (following the procedure of ref. 26). The  $P$  values at each bootstrap iteration were ranked, and for each coefficient, the median and width of the 95% confidence interval of the ranks across the 250 bootstrap samples were calculated. For each pairwise model, the  $i$ th metabolic variable was then considered discriminatory if, separately for both first and second 24-h urinary specimens, the following criteria were met: (1) the coefficients  $b_i$  were significant,  $P_i < 7 \times 10^{-6}$  (corresponding to  $P < 0.05$  after Bonferroni correction for 7,100 spectral variables; this is conservative given the correlation structure between spectral variables); (2) the

median ranks for the  $i$ th coefficient were in the top 5%; (3) the width of the confidence intervals of the ranks were in the bottom 5%; (4) the coefficients  $b_i$  were in the top 60%.

**Quantitation of metabolites, reliability and regression against blood pressure.** Mean concentrations of four metabolites (alanine, formate, hippurate and  $N$ -methylnicotinate) were quantified from the NMR spectra of first and second urine specimens. We used an automated method modified from ref. 27, and excluded specimens where the estimation procedure failed or the values fell outside the method tolerance limits, for either urine specimen. Results were calibrated to the creatinine peak ( $\text{CH}_2$   $\delta = 4.06$ ) and then to creatinine concentration measured externally using the Jaffé method (Supplementary Information). We calculated 24-h urinary excretion (mmol per 24 h) by multiplying urinary concentrations by urinary volume. Technical error<sup>8</sup> for each quantified metabolite was calculated from the 8% split specimens. To compare the NMR findings with independent analyses, we calculated the Spearman rank correlation coefficients relating our results (first urine specimens) for alanine ( $n = 4,232$ ) with those from ion exchange chromatography<sup>28</sup>, and, for hippurate ( $n = 124$ ), with gas chromatography mass spectrometry<sup>29</sup>.

Observed regression coefficients relating urinary variables to BP are attenuated because of within-person variability<sup>24</sup> in metabolite excretion. We estimated within- and between-person variance for urinary metabolites from one-way ANOVA. Ratios of within- to between-person variance,  $\lambda$ , were calculated for eight country/gender-specific subgroups and pooled, weighted by degrees of freedom. We averaged the two 24-h urinary values for each quantified metabolite, and estimated percentage of the theoretical regression coefficient ( $K_{xx}$ ) for univariate regression by the formula  $K_{xx} = 2/(2 + \lambda) \times 100$  (ref. 24). We used multiple regression to relate mean metabolite concentrations to mean systolic and diastolic BP using SAS (version 9.1, SAS Institute) with adjustment for potential confounders, with and without body mass index<sup>6</sup>. We fitted regression models by country and pooled coefficients across countries, weighted by inverse of the variance, to estimate overall association, and tested for heterogeneity of country-specific coefficients<sup>6</sup>. We express regression coefficients as mm Hg per 2 s.d. higher urinary metabolite excretion, from pooled within-country standard deviations (from one-way ANOVA)<sup>6</sup>.

**Structural characterization of metabolites.** We used available spectral databases and chemical addition experiments to aid structural identification of discriminatory metabolites. For the remaining metabolites, we used statistical total correlation spectroscopy<sup>30</sup> and solid phase extraction chromatography coupled with NMR<sup>29</sup>.

25. Wold, S. Cross-validatory estimation of number of components in factor and principal components models. *Technometrics* **20**, 397–405 (1978).
26. Martens, H. & Martens, M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Qual. Prefer.* **11**, 5–16 (2000).
27. Crockford, D. J. et al. Curve fitting method for direct quantitation of compounds in complex biological mixtures using  $^1\text{H}$  NMR: Application in metabonomic toxicology studies. *Anal. Chem.* **77**, 4556–4562 (2005).
28. Fekkes, D., Voskuilen-Kooyman, A., Jankie, R. & Huijman, J. Precise analysis of primary amino acids in urine by an automated high-performance liquid chromatography method: Comparison with ion-exchange chromatography. *J. Chromatogr. B* **744**, 183–188 (2000).
29. Lenz, E. M. & Wilson, I. D. Analytical strategies in metabonomics. *J. Proteome Res.* **4**, 443–458 (2007).
30. Cloarec, O. et al. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic  $^1\text{H}$  NMR data sets. *Anal. Chem.* **77**, 1282–1289 (2005).



# Genetic evidence that FGFs have an instructive role in limb proximal–distal patterning

Francesca V. Mariani<sup>1\*†</sup>, Christina P. Ahn<sup>1\*</sup> & Gail R. Martin<sup>1</sup>

Half a century ago, the apical ectodermal ridge (AER) at the distal tip of the tetrapod limb bud was shown to produce signals necessary for development along the proximal–distal (P–D) axis, but how these signals influence limb patterning is still much debated<sup>1,2</sup>. Fibroblast growth factor (FGF) gene family members are key AER-derived signals<sup>3,4</sup>, with *Fgf4*, *Fgf8*, *Fgf9* and *Fgf17* expressed specifically in the mouse AER<sup>5</sup>. Here we demonstrate that mouse limbs lacking *Fgf4*, *Fgf9* and *Fgf17* have normal skeletal pattern, indicating that *Fgf8* is sufficient among AER-FGFs to sustain normal limb formation. Inactivation of *Fgf8* alone causes a mild skeletal phenotype<sup>6,7</sup>; however, when we also removed different combinations of the other AER-FGF genes, we obtained unexpected skeletal phenotypes of increasing severity, reflecting the contribution that each FGF can make to the total AER-FGF signal. Analysis of the compound mutant limb buds revealed that, in addition to sustaining cell survival, AER-FGFs regulate P–D patterning gene expression during early limb bud development, providing genetic evidence that AER-FGFs function to specify a distal domain and challenging the long-standing hypothesis that AER-FGF signalling is permissive rather than instructive for limb patterning. We discuss how a two-signal model for P–D patterning can be integrated with the concept of early specification to explain the genetic data presented here.

*Fgf8* is expressed in prospective AER cells of the nascent limb bud and, subsequently, throughout the AER until it regresses<sup>8</sup>. By contrast, *Fgf4*, *Fgf9* and *Fgf17* expression commences after the AER is formed, is restricted to the posterior AER, and ceases at least a day before AER regression<sup>5</sup> (Fig. 1a). When AER-FGFs are individually eliminated, only loss of *Fgf8* function perturbs skeletal patterning<sup>5–7,9–11</sup>. The other AER-FGFs have been proposed to be essential, but functionally redundant, components of a positive-feedback loop between the AER and the patterning centre in posterior limb bud mesenchyme that produces sonic hedgehog (SHH)<sup>5,12,13</sup>. We tested this hypothesis by deleting *Fgf4* by means of Cre-mediated recombination in the AER of embryos homozygous for *Fgf9* and *Fgf17* null alleles<sup>10,11</sup> (hereafter referred to as F4;9,17-triple knockout (TKO) mutants; Fig. 1b). Because *Fgf4* deletion occurs before *Fgf4* expression normally commences<sup>5</sup> (see Fig. 1a), the F4;9,17-TKO limb buds do not produce FGF4, FGF9 or FGF17. Nevertheless, in F4;9,17-TKO skeletons ( $n = 6$ ), the three classically defined limb segments—stylopod (upper arm or leg), zeugopod (lower arm or leg) and autopod (wrist/hand or ankle/foot)—were essentially normally patterned (Fig. 1c). Consistent with this observation, *in situ* hybridization (not shown) and quantitative PCR after reverse transcription (qRT–PCR; Fig. 1d) showed normal *Shh* expression in F4;9,17-TKO limb buds at embryonic day (E)10.5. Moreover, there was no compensatory upregulation of *Fgf8* in F4;9,17-TKO limb buds at E10.5 (Fig. 1d). These data demonstrate that *Fgf8* is sufficient for

normal limb development, including sustaining *Shh* expression, and that whatever positive regulatory interactions occur between *Shh* and the posterior AER-FGF genes, they are dispensable for normal limb skeletal development.

Although not required when FGF8 is present, each posterior AER-FGF (FGF4, FGF9 and FGF17) may contribute to limb development. Such contributions can be uncovered by inactivating these genes, singly or in combination, along with *Fgf8* (refs 14 and 15). To produce such double-knockout (DKO) and TKO mutants, we used *Msx2-cre*, which functions earlier in hindlimb than in forelimb buds<sup>5</sup> (see Fig. 1a). When inactivated by *Msx2-cre*, *Fgf8* is never expressed in hindlimb buds, but is transiently expressed in forelimb buds before E9.5; consequently, *Fgf8* knockout (F8-KO) hindlimbs are more severely affected than forelimbs<sup>6</sup>. In F8;4-DKO mutants, the hindlimb skeleton fails to form, whereas the forelimb skeleton develops but lacks many elements<sup>14</sup>. Likewise, in F8;4-DKO mutants lacking a copy of *Fgf9* (F8;4-DKO;*Fgf9*<sup>+/−</sup> mutants), there was no hindlimb, but a more severely affected forelimb developed (see below). Compound mutant forelimbs thus provided a greater range of phenotypes for analysis than hindlimbs.

These forelimb phenotypes (summarized in Supplementary Table 1) could be ranked in order of increasing severity. F8-KO ( $n = 10$ ), F8;17-DKO ( $n = 8$ ) and F8-KO;*Fgf9*<sup>+/−</sup> ( $n = 6$ ) mutants displayed a similar mild phenotype, with all skeletal elements present except for one digit, and slight hypoplasia of the stylopod and zeugopod (Fig. 2a, b; data not shown; ref. 6). Similar defects were seen in F8;9-DKO ( $n = 16$ ) and F8;9,17-TKO ( $n = 8$ ) mutants, but the zeugopod posterior element (ulna) was short and the anterior element (radius) was absent (Fig. 2c; data not shown). The hindlimb zeugopod was similarly affected in these mutants, but, notably, it was the posterior zeugopod element (fibula) that was missing (not shown). Assays for *Sox9* expression, which marks the condensations that develop into skeletal elements<sup>16</sup>, showed that these patterning defects were detectable by E12.5 (Fig. 2h–j; data not shown), demonstrating that AER-FGFs are essential for establishing skeletal pattern at limb bud stages. A more severe phenotype was observed in F8;4-DKO mutants ( $n = 5$ ), in which the forelimb zeugopod consisted of only a hypoplastic ulna and all autopod elements were absent except for one or two phalanges (Fig. 2d; ref. 14). Removing one copy of *Fgf9* further increased phenotype severity ( $n = 14$  mutants; Fig. 2e, f); when both copies were removed, F8;4;9-TKO mutants ( $n = 9$ ) lacked all forelimb skeletal elements (Fig. 2g).

The simplest explanation for these data is that the individual AER-FGFs are functionally equivalent<sup>17</sup> but that they differ in the extent to which they contribute to the AER-FGF signal, presumably reflecting differences in their temporal and spatial expression profiles, levels of expression, and binding specificities to FGF receptors in the limb bud mesenchyme. If so, then the range of skeletal phenotypes observed

<sup>1</sup>Department of Anatomy and Program in Developmental Biology, School of Medicine, University of California at San Francisco, San Francisco, California 94158-2324, USA. <sup>†</sup>Present address: Broad Center for Stem Cell Research and Regenerative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.

\*These authors contributed equally to this work.

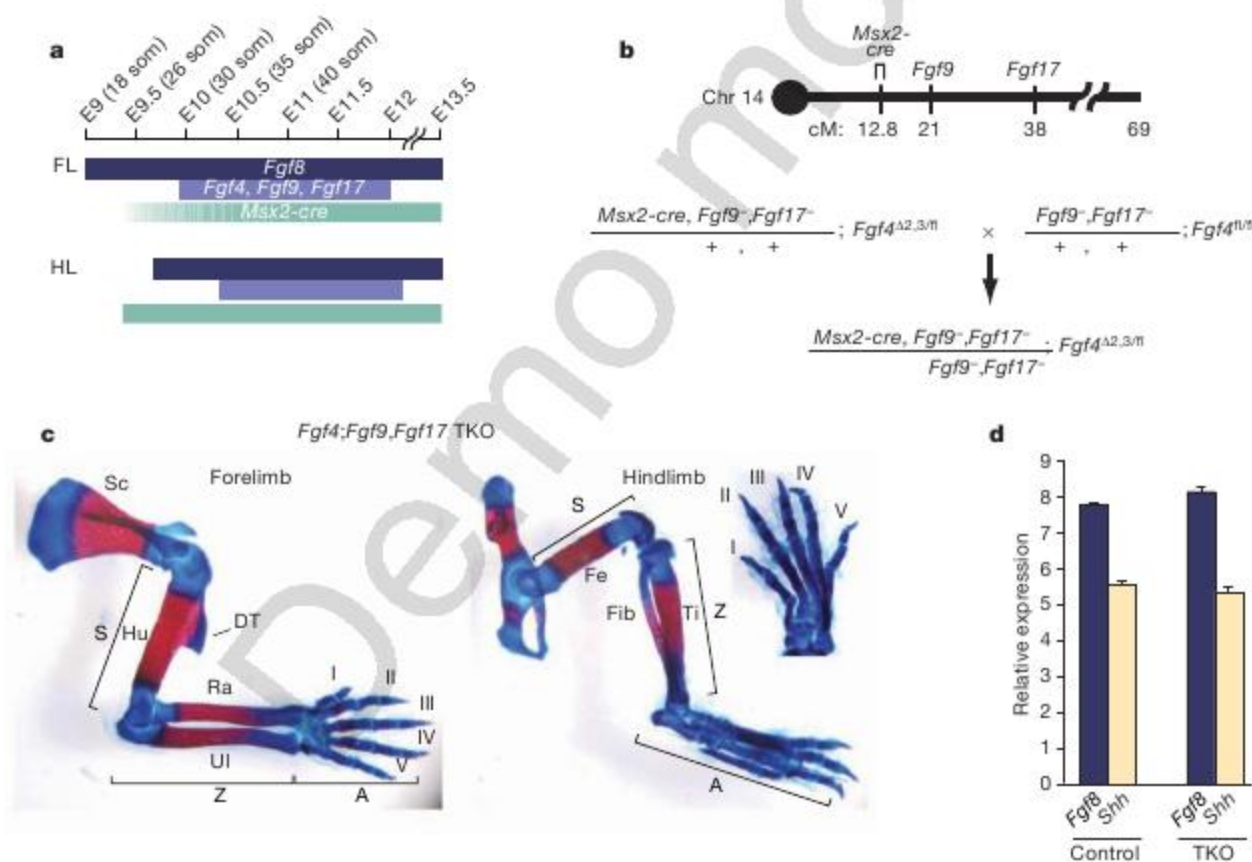


when specific combinations of AER-FGFs are removed reflects a change in the level of total AER-FGF signal. In support of this hypothesis, we found that at E10.5 the size and intensity of the expression domain of *Dusp6* (a downstream target of AER-FGF signalling<sup>18</sup>) negatively correlated with the severity of the mutant phenotype (Fig. 2l–o). Together, these data suggest that *Fgf8* makes the greatest contribution to the AER-FGF signal, followed by *Fgf4*, *Fgf9* and *Fgf17*. Furthermore, our data suggest that there is a critical threshold of AER-FGF signalling, below which skeletal elements do not form.

The phenotype of the *F8;4-DKO;F9<sup>-/+</sup>* forelimbs was especially notable: the stylopod (humerus) was present, but was often smaller than normal and truncated distally, with elements of a single digit immediately distal to it. The most distal element usually had the pointed tip of a terminal phalanx, occasionally with a nail overlying it (Fig. 2e, f; data not shown). Sometimes there was a substantial gap between the humerus and the phalangeal elements (Fig. 2f and Supplementary Table 1). The complete absence of the zeugopod and most of the autopod was confirmed by *Sox9* expression assays at E12.5 (Fig. 2k; see Supplementary Fig. 1 for data on *Hoxa11* expression). One possible explanation for this phenotype is that the severe reduction in AER-FGF signalling causes the death of most autopod and zeugopod progenitors in the early limb bud. However, consistent with previous reports<sup>14,15</sup>, mesenchymal cell death in *F8;4-DKO;F9<sup>-/+</sup>* mutants was detected in a proximal dorsal region at

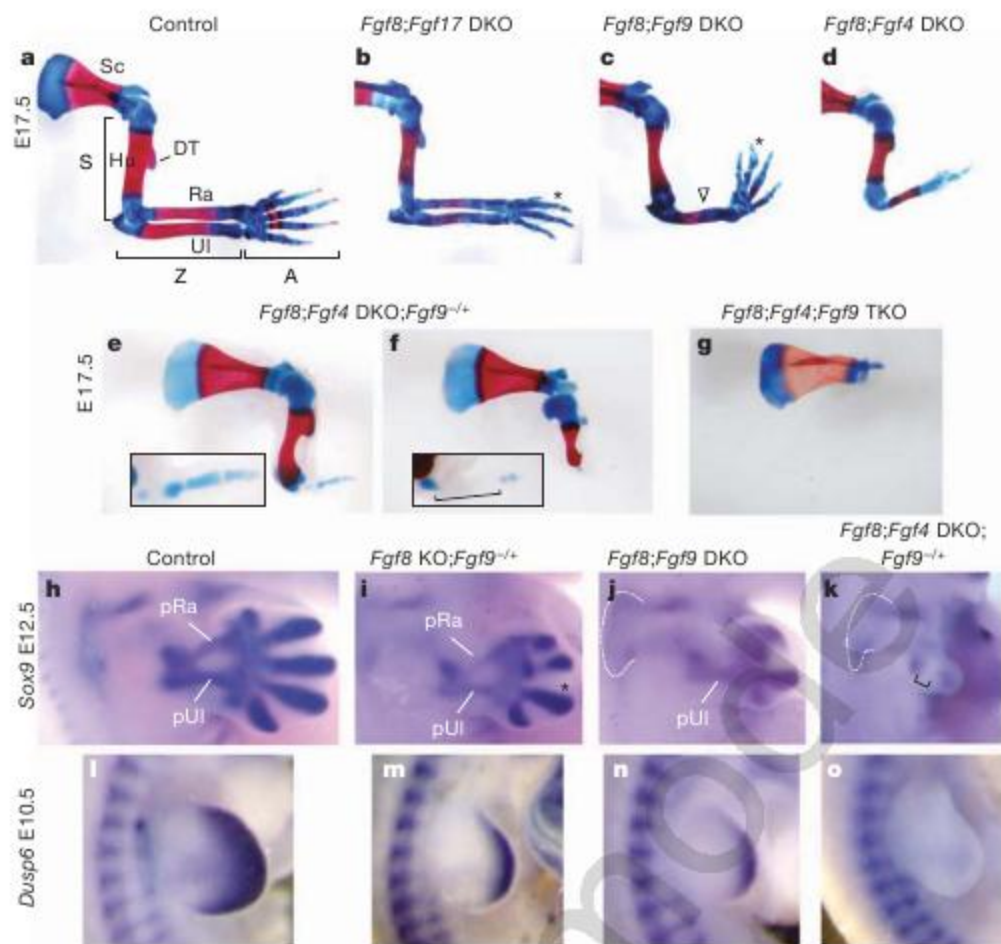
E10.5, and not distally (Fig. 3c), where autopod and distal zeugopod progenitors reside<sup>19</sup>. Cell death remained proximally restricted until E11.5, after which it was no longer detected (not shown). In mutant limb buds with more AER-FGF signalling, the cell death domain occupied a smaller percentage of limb bud volume (Fig. 3a–d; data not shown). Because the dying cells are proximally localized, it seems unlikely that autopod or zeugopod progenitor cell death is the main cause of the phenotype.

Another possible explanation for the lack of zeugopod and most of the autopod in *F8;4-DKO;F9<sup>-/+</sup>* forelimbs, is that AER-FGFs are involved in specifying distal cell fate, and when AER-FGF signalling is markedly reduced, fewer cells are specified as distal. This hypothesis could be tested by examining the effects of reducing AER-FGF signalling on the expression of genes essential for P–D specification. *Meis1*, which encodes a homeobox transcription factor, is potentially one such gene because ectopic *Meis1* expression is reported to induce distal-to-proximal transformations in chicken limb buds<sup>20</sup>. Initially, *Meis1* expression is detected throughout the nascent limb bud mesenchyme; subsequently, a *Meis1*-negative distal domain is established and increases in size as the limb bud expands<sup>20</sup>. Consistent with a role for FGFs in specification of a distal domain, FGF-bead implantation and drug inhibition studies showed that FGF signalling can repress *Meis1* expression<sup>21</sup>. Moreover, we previously found that the extent of the *Meis1*-negative expression domain was reduced in



**Figure 1** | *Fgf8* is sufficient for normal limb development. **a**, Schematic diagram illustrating temporal aspects of wild-type AER-FGF gene expression and the stages at which the *Msx2-cre* transgene functions to inactivate the *Fgf8* and *Fgf4* floxed alleles in the AER. Note that development of forelimb (FL) buds, as marked by *Fgf8* expression, commences before that of hindlimb (HL) buds, that *Fgf8* expression precedes that of *Fgf4*, *Fgf9* and *Fgf17*, and that *Msx2-cre* functions earlier in hindlimb than in forelimb buds<sup>3,6</sup>. **b**, Schematic diagram of mouse chromosome 14, showing the map positions of *Fgf9*, *Fgf17* and the *Msx2-cre* transgene, which lies within 1 centimorgan (cM) of *Bmpr1a* (not illustrated) and 12.8 cM from the centromere (circle)<sup>28</sup>. Because of this linkage, once the parental animals (male on left; female on right) were generated we could produce progeny of the genotype illustrated (*Fgf4;Fgf9;Fgf17* TKO mutants), in which the *Fgf4* conditional null allele (*Fgf4<sup>Δ1</sup>*) is converted by *Msx2-cre* function in the AER to *Fgf4<sup>Δ2,3</sup>*, a null allele lacking exons 2 and 3, at a frequency of 12.5% ( $n = 6/48$ ), close to the expected frequency of 15.5%. **c**, *Fgf4;Fgf9;Fgf17* TKO forelimb and hindlimb skeletons at E17.5; these are indistinguishable from those of the wild type (not shown), except for an enlarged deltoid tuberosity caused by loss of *Fgf9* function after condensation<sup>29</sup>. **d**, Quantitative RT-PCR assays for *Fgf8* and *Shh* expression. A representative experiment on forelimb buds from embryos at ~E11.0 (39–40 somites;  $n = 4$  limb buds for each genotype) is shown. Values are normalized to cyclophilin expression and are shown as means  $\pm$  standard deviation. The difference between control and mutant limb buds with respect to *Fgf8* and *Shh* expression was not significant (*Fgf8*,  $P = 0.61$ ; *Shh*,  $P = 0.80$ ). A two-tailed Student's *t*-test was employed, using the average of triplicate cycle count values for each limb bud. Similar results were obtained for limb buds at 33, 34, 35 and 37 somites. Abbreviations: A, autopod; DT, deltoid tuberosity; Fe, femur; Fib, fibula; Hu, humerus; Ra, radius; Sc, scapula; S, stylopod; som, somite number; Ti, tibia; Ul, ulna; Z, zeugopod; I–V, digit numbers from anterior to posterior.





**Figure 2 | Effects of inactivating AER-FGF genes on skeletal development.** **a–g**, Comparison of skeletal preparations of forelimbs from E17.5 embryos of the genotypes indicated. The asterisk in **b**, **c** and **i** indicates that the mutant autopod has only four digits. The open down-triangle (**c**) indicates the lack of the anterior element (radius). The differences in humerus thickness/shape and deltoid tuberosity size among mutants of the various genotypes illustrated in **b–d** were also observed among mutants of the individual genotypes, suggesting they are caused by background genes. **e**, **f**, Two examples of *Fgf8;Fgf4*-DKO;*Fgf9*<sup>-/-</sup> forelimb skeletons, illustrating the more (**e**) and less (**f**) common phenotypes, respectively. The insets show the distal element of the limbs at higher magnification. The bracket (**f**) indicates

the gap between the distal end of the humerus and the digit-like element. **h–o**, Expression of *Sox9* at E12.5 and of *Dusp6* at E10.5 (37 somites), as detected by RNA *in situ* hybridization in whole-mount in forelimb buds from embryos of the genotypes indicated. *Sox9* expression marks the condensations that prefigure the skeletal elements. The dotted white lines (**j** and **k**) outline the condensations that will develop into the scapula. Note the absence of the developing radius (**j**). The bracket (**k**) demarcates the region devoid of *Sox9*-positive cells between the developing condylar portion of the humerus and a distal condensation that presumably represents the distal part of a digit. pRa, prospective radius; pUI, prospective ulna. Other abbreviations are as in the legend to Fig. 1.

*F8;F4*-DKO limb buds<sup>14</sup>. However, because the *F8;F4*-DKO limb buds were considerably smaller than those of controls, it was impossible to know whether the observed effect was secondary to the reduction in limb bud size.

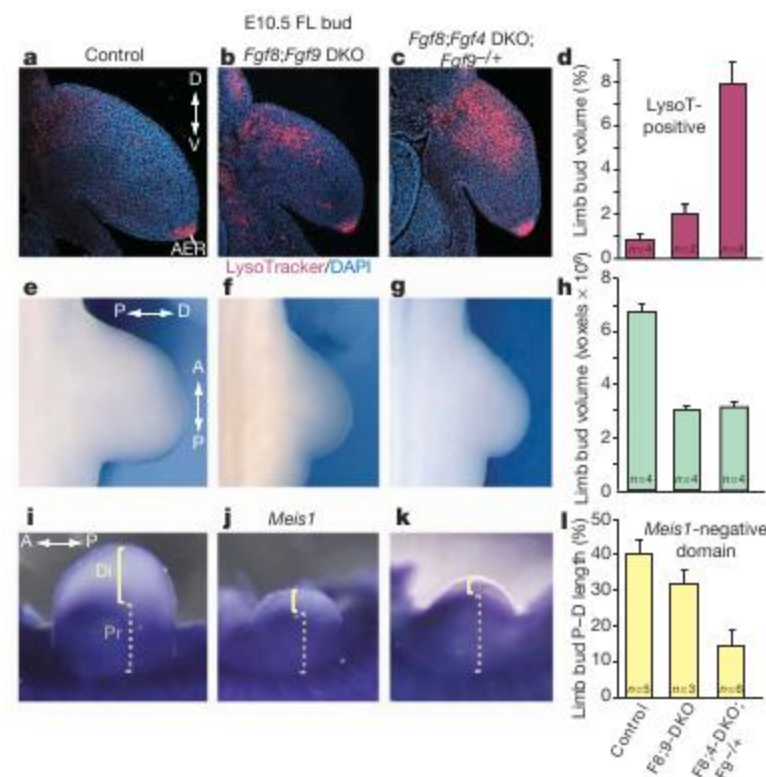
We were able to examine the effects of reducing AER-FGF signalling on *Meis1* expression independent of its effects on limb bud size, because the various AER-FGF compound mutant limb buds, although smaller than normal, were remarkably similar in overall size at E10.5 (Fig. 3e–h; data not shown). One explanation for this finding is that their size at E10.5 reflects the elimination of cells that died due to inactivation of *Fgf8*, which is expressed from ~E9.0 and inactivated by ~E9.5, whereas the effects of loss of *Fgf4* and *Fgf9* function on cell survival are not yet evident because their expression begins later than *Fgf8* expression. By E11.5, however, size differences became evident among the compound mutants; this was caused by extensive cell death and possibly a negative effect on cell proliferation after E10.5. These size differences were correlated with the decrease in AER-FGF signalling (not shown) and, subsequently, with skeletal phenotype severity (Fig. 2h–k and data not shown). Assays for *Meis1* at E10.5, when limb bud size was similar, showed that the *Meis1*-negative, distal domain was significantly reduced in *F8;9*-DKO and was further reduced in *F8;4*-DKO;*F9*<sup>-/-</sup> forelimb buds (Fig. 3i–l). To our knowledge, these data provide the first genetic evidence that AER-FGFs repress expression of a gene presumed to be involved in specification of proximal cell

identity. In summary, we conclude that AER-FGF signalling serves at least two vital roles during limb development: to promote cell survival and to specify distal cell fate.

An important question is how our data fit with existing models for limb P–D patterning. The ‘progress zone’ model<sup>22</sup> postulates that the P–D pattern develops gradually, with cells in the distal limb bud mesoderm acquiring progressively more distal positional information over time, and that AER signals are not instructive for P–D patterning, but instead are ‘permissive’, keeping distal cells labile and able to change positional values by an unknown mechanism. However, our data showing that AER-FGF signalling is necessary to regulate *Meis1* expression indicate that it functions as an instructive rather than a permissive signal in limb development. Furthermore, the phenotypes of AER-FGF mutant skeletons (for example, *F8;4*-DKO;*F9*<sup>-/-</sup> forelimbs, which contain stylopod and autopod but no zeugopod elements) are not readily explicable by the progress zone model<sup>1,14</sup>, which describes limb patterning as a progressive process whereby distal is specified only after proximal.

By contrast, the ‘early specification’ model<sup>23</sup> postulates that cells along the limb P–D axis are specified to form the stylopod, zeugopod and autopod limb segments at an early limb bud stage. Although not originally considered in this context, one model that can explain how such early specification occurs is the ‘two-signal model’. This proposes that limb bud cells are initially exposed to a proximal signal



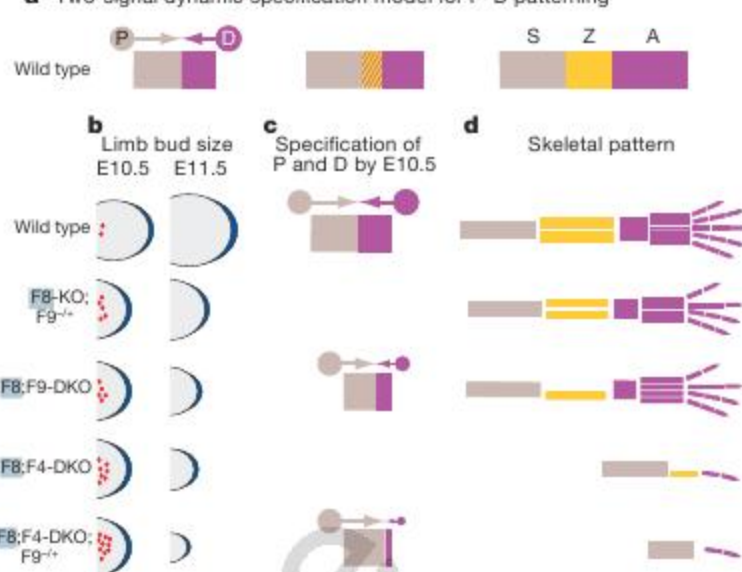


**Figure 3 | Effects of inactivating AER-FGF genes on cell survival, limb bud size and *Meis1* expression.** **a–c**, Confocal images of sections through E10.5 forelimb limb buds (35–36 somites) of the genotypes indicated stained with LysoTracker Red (which labels apoptotic cells as well as healthy cells engulfing apoptotic debris<sup>30</sup>) and DAPI (blue, which labels nuclei). **d**, Graph showing the average percentage of limb bud volume that is LysoTracker-Red-positive for each genotype. The difference between the two mutants was statistically significant ( $P = 0.014$ , two-tailed Student's *t*-test). **e–g**, Dorsal views of E10.5 forelimb buds (36–37 somites) from embryos of the genotypes indicated. **h**, Graph showing the average total volume of E10.5 limb buds (35–37 somites) of the genotypes indicated. No statistically significant difference was detected between mutants ( $P = 0.79$ , two-tailed Student's *t*-test). **i–k**, E10.5 forelimb buds (37 somites) were assayed in whole mount by RNA *in situ* hybridization for *Meis1* expression. The dashed tan and solid yellow brackets in each panel demarcate the *Meis1*-positive proximal and *Meis1*-negative distal domains in the limb bud mesenchyme, respectively. The length of each bracket reflects the average of the four measurements made on each sample. **l**, The average percentage of total limb bud proximal–distal (P–D) length (the sum of the lengths of the proximal and distal domains) that is *Meis1*-negative is shown for each limb bud. The difference between mutants was statistically significant ( $P = 0.0003$ , two-tailed Student's *t*-test). In **d**, **h** and **l**, *n* defines the number of limb buds analysed for each genotype, and the error bars show the standard deviation. Abbreviations: A ↔ P, anterior–posterior; AER, apical ectodermal ridge; D ↔ V, dorsal–ventral; P ↔ D, proximal–distal.

from mesoderm flanking the limb bud, possibly retinoic acid, and then to an opposing distal signal (FGF) from the AER, which establish proximal and distal domains, respectively<sup>2,21</sup>. Formation of a third (middle) domain might then occur as a result of interactions between cells at the boundary between proximal and distal domains over time, thus creating the three domains from which the stylopod, zeugopod and autopod segments will develop (see Fig. 4a). Additional domains within the autopod, from which wrist elements, metacarpals and digits develop, might likewise form as a result of cell–cell interactions at domain boundaries. This concept is consistent with the intercalation models proposed to explain amphibian limb regeneration<sup>24,25</sup>. Early specification would thus be a ‘dynamic’ process that takes place in concert with limb bud outgrowth. Our data support this model by providing genetic evidence that AER-FGFs function as an initial distal signal at early stages in limb development.

Such a two-signal dynamic specification model can easily explain the forelimb skeletal abnormalities reported here based on our

**a** Two-signal dynamic specification model for P–D patterning



**Figure 4 | AER-FGF mutant phenotypes can be explained by a two-signal dynamic specification model for limb proximal–distal patterning.** **a**, Left, proximal (P) signal(s) from flanking mesoderm and opposing distal (D) signal(s) from the AER (FGFs) specify P (tan) and D (purple) domains<sup>2,21</sup> in the limb bud by E10.5; centre, a third, middle domain (cross-hatched) forms within the D domain, a hypothesis supported by recent fate mapping studies<sup>19</sup>; right, the three domains develop into the classically defined segments: stylopod (S, tan), zeugopod (Z, gold) and autopod (A, purple). **b**, Wild-type and AER-FGF mutant forelimb buds (dorsal views). At E10.5, all mutant limb buds are equal in size but are smaller than the wild type (quantified in this study for all genotypes except F8;4-DKO), and contain dying cells in the mesenchyme (red dots; cell death in AER not depicted). By E11.5, mutant limb bud size is decreased in proportion to the reduction in AER-FGF signal; dying cells are no longer observed in the mesenchyme. **c**, *Meis1*-positive P (tan) and *Meis1*-negative D (purple) domains are specified by opposing P and D signals, respectively. In the mutants, only the D signal is reduced, in proportion to the AER-FGF signal. Because the D signal is weaker than normal and also because the limb bud is smaller than normal due to cell death specifically in the P domain, the P signal extends over a region that, in wild-type limb buds, would normally be exposed to the D signal. **d**, Diagrams representing the size and number of elements in forelimbs of each genotype at birth. Owing to changes in the extent of limb bud P–D length influenced by P and D signals in the mutants at E10.5 (arrows in **c**), cells normally fated to form Z or A elements are specified as proximal and instead develop into S. For example, in F8;9-DKO limb buds, some Z and A elements are missing whereas S is nearly normal. In F8;4-DKO;F9<sup>-/-</sup> limb buds, all Z and all but the distal-most A elements are missing, whereas S forms but is reduced in size.

findings that AER-FGFs have a dual function in limb development. Thus, we suggest that in the AER-FGF mutants, the distal signal is reduced in proportion to the decrease in AER-FGF signal. Because there is less opposing distal signal, the proximal signal, which is produced at the normal level, extends more distally than normal, specifying cells that would normally have formed distal elements to be proximal and thereby compromising autopod and zeugopod development. However, the stylopod is not abnormally long because proximal cells die in the AER-FGF mutants (Fig. 4b–d; see Supplementary Fig. 2a, b). This model of AER-FGF dual function can also explain the limb skeletal defects observed in other studies of mouse mutants<sup>6,7,34</sup> and X-irradiated chicken limb buds<sup>26</sup> in which the stylopod is severely reduced but distal elements are less affected (see Supplementary Fig. 2c).

Presently, it is not possible to assess when segment specification occurs and critical tests of this model cannot be performed because there are no molecular markers for the progenitors of the different segments. A promising avenue for identifying such markers has recently been opened by the demonstration that cells in proximal and distal regions of the early chicken limb bud are distinguishable



by their sorting behaviours *in vitro*<sup>27</sup>. It will be intriguing to determine whether FGFs produced in the AER are involved in establishing these differences.

## METHODS SUMMARY

**Production of mice.** The mutant alleles used in this study were maintained on a mixed genetic background. F4;9,17-TKO mutants were generated by crossing parental animals of the genotypes depicted in Fig. 1b. Animals of the other genotypes were produced by generating and crossing appropriate parental mice. Control animals were offspring that did not inherit the *Msx2-cre* transgene.

**Limb bud analysis.** For qRT-PCR, complementary DNA was synthesized and analysed using an ABI light cycler. For analysis of cell death, embryos were stained with LysoTracker Red (Invitrogen) and sectioned. For each section, the LysoTracker-Red-positive and limb bud areas were measured using tools in Adobe Photoshop. The *Meis1*-negative domain was determined for each sample by four independent, blind measurements that were then averaged.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 November 2007; accepted 22 February 2008.

Published online 30 April 2008.

- Mariani, F. V. & Martin, G. R. Deciphering skeletal patterning: clues from the limb. *Nature* **423**, 319–325 (2003).
- Tabin, C. & Wolpert, L. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev.* **21**, 1433–1442 (2007).
- Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* **75**, 579–587 (1993).
- Fallon, J. F. et al. FGF-2: apical ectodermal ridge growth signal for chick limb development. *Science* **264**, 104–107 (1994).
- Sun, X. et al. Conditional inactivation of *Fgf4* reveals complexity of signalling during limb bud development. *Nature Genet.* **25**, 83–86 (2000).
- Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
- Moon, A. M. & Capecchi, M. R. Fgf8 is required for outgrowth and patterning of the limbs. *Nature Genet.* **26**, 455–459 (2000).
- Crossley, P. H. & Martin, G. R. The mouse *Fgf8* gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439–451 (1995).
- Moon, A. M., Boulet, A. M. & Capecchi, M. R. Normal limb development in conditional mutants of *Fgf4*. *Development* **127**, 989–996 (2000).
- Colvin, J. S., White, A. C., Pratt, S. J. & Ornitz, D. M. Lung hypoplasia and neonatal death in *Fgf9*-null mice identify this gene as an essential regulator of lung mesenchyme. *Development* **128**, 2095–2106 (2001).
- Xu, J., Liu, Z. & Ornitz, D. M. Temporal and spatial gradients of *Fgf8* and *Fgf17* regulate proliferation and differentiation of midline cerebellar structures. *Development* **127**, 1833–1843 (2000).
- Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. *Sonic hedgehog* and *Fgf-4* act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993–1003 (1994).
- Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609–612 (1994).
- Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–508 (2002).
- Boulet, A. M., Moon, A. M., Arenkiel, B. R. & Capecchi, M. R. The roles of *Fgf4* and *Fgf8* in limb bud initiation and outgrowth. *Dev. Biol.* **273**, 361–372 (2004).
- Wright, E. et al. The *Sry*-related gene *Sox9* is expressed during chondrogenesis in mouse embryos. *Nature Genet.* **9**, 15–20 (1995).
- Lu, P., Minowada, G. & Martin, G. R. Increasing *Fgf4* expression in the mouse limb bud causes polysyndactyly and rescues the skeletal defects that result from loss of *Fgf8* function. *Development* **133**, 33–42 (2006).
- Kawakami, Y. et al. MKP3 mediates the cellular response to FGF8 signalling in the vertebrate limb. *Nature Cell Biol.* **5**, 513–519 (2003).
- Sato, K., Koizumi, Y., Takahashi, M., Kuroiwa, A. & Tamura, K. Specification of cell fate along the proximal–distal axis in the developing chick limb bud. *Development* **134**, 1397–1406 (2007).
- Mercader, N. et al. Conserved regulation of proximodistal limb axis development by *Meis1/Hth*. *Nature* **402**, 425–429 (1999).
- Mercader, N. et al. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of *Meis* genes. *Development* **127**, 3961–3970 (2000).
- Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
- Dudley, A. T., Ros, M. A. & Tabin, C. J. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* **418**, 539–544 (2002).
- Maden, M. Intercalary regeneration in the amphibian limb and the rule of distal transformation. *J. Embryol. Exp. Morphol.* **56**, 201–209 (1980).
- Stocum, D. L. The urodele limb regeneration blastema. Determination and organization of the morphogenetic field. *Differentiation* **27**, 13–28 (1984).
- Wolpert, L., Tickle, C. & Sampford, M. The effect of cell killing by x-irradiation on pattern formation in the chick limb. *J. Embryol. Exp. Morphol.* **50**, 175–193 (1979).
- Barna, M. & Niswander, L. Visualization of cartilage formation: insight into cellular properties of skeletal progenitors and chondrodysplasia syndromes. *Dev. Cell* **12**, 931–941 (2007).
- Pajni-Underwood, S., Wilson, C. P., Elder, C., Mishina, Y. & Lewandoski, M. BMP signals control limb bud interdigital programmed cell death by regulating FGF signaling. *Development* **134**, 2359–2368 (2007).
- Hung, I. H., Yu, K., Lavine, K. J. & Ornitz, D. M. FGF9 regulates early hypertrophic chondrocyte differentiation and skeletal vascularization in the developing stylopod. *Dev. Biol.* **307**, 300–313 (2007).
- Zucker, R. M., Hunter, E. S. III & Rogers, J. M. Apoptosis and morphology in mouse embryos by confocal laser scanning microscopy. *Methods* **18**, 473–480 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Ornitz for providing mice carrying the *Fgf9* and *Fgf17* null alleles before publication, and D. Lakeland for help with the statistical analysis. We are grateful to P. Ghatpande, A. Nemati and E. Yu for technical assistance. We also thank M. Barna, X. Sun, R. Metzger and our laboratory colleagues for comments on the manuscript. F.M. was supported by a fellowship from the American Heart Association and a National Research Service Award (NIH). This work was supported by a grant from the National Institutes of Health (R01 HD34380) to G.R.M.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.R.M. ([gail.r.martin@ucsf.edu](mailto:gail.r.martin@ucsf.edu)).



## METHODS

**Production and analysis of mice.** Strategies similar to that used for generating the F4;9,17-TKO mutants (Fig. 1b) were used to generate animals of the other genotypes used in this study. For example, to generate F8;9-DKO or F8-KO;F9<sup>+/−</sup> offspring, *Msx2-cre*;Fgf9<sup>+/−</sup>;Fgf8<sup>Δ2,3/Δ</sup> males were crossed with Fgf9<sup>+/−</sup>;Fgf8<sup>Δ/Δ</sup> or Fgf8<sup>Δ/Δ</sup> females. F4;8;9-TKO or F4;8-DKO;F9<sup>+/−</sup> animals were generated by crossing *Msx2-cre*;Fgf9<sup>+/−</sup>;Fgf4<sup>Δ2,3/Δ</sup>;Fgf8<sup>Δ/Δ</sup> males with Fgf9<sup>+/−</sup>;Fgf4<sup>Δ/Δ</sup>;Fgf8<sup>Δ/Δ</sup> or Fgf4<sup>Δ/Δ</sup>;Fgf8<sup>Δ/Δ</sup> females. F8-KO and F8;F4-DKO animals were generated as previously described<sup>6,14</sup>. Animals were genotyped by PCR on DNA isolated from head or tail tissue, using primer pairs for *Msx2-cre*, *Fgf4* and *Fgf8* alleles<sup>14</sup>. For *Fgf9*, we used the following primers: wild-type allele forward, 5'-GCAAGGGAGGGG-AGTTGGATATACC-3', and reverse, 5'-GAAATCCAGTCCTGCAGTACAG-CTGC-3'; mutant allele forward, 5'-CCTTTTCCCTCTCTGTCTGCAAC-3', and reverse, 5'-TGTGCTCTAGTAGCTTTACGGAGCC-3'. For *Fgf17*, the primers were: wild-type allele forward, 5'-GAAGTTTCTCCAGCGATGGG-3', and reverse, 5'-GACAGCAGAGAATCAATAGCTGC-3'; mutant allele forward, 5'-CCATGAGTGAACGAACCTGG-3', and reverse, 5'-TTGGCTTCTCTGGGACT-CTACC-3'. Because homozygosity for *Fgf9* function causes perinatal lethality<sup>10</sup>, we assayed for skeletal pattern at E17.5 by staining cartilage and bone with alcian blue and alizarin red, respectively, using standard procedures. Embryos younger than E11 were staged by counting somites as described previously<sup>5</sup>. Whole-mount RNA *in situ* hybridization was performed according to a standard protocol.

**Limb bud analysis.** Limbs for qRT-PCR were dissected and stored at −20 °C in RNAlater (Qiagen, catalogue number 1017980) until genotyping information was obtained. Individual limbs were then homogenized with the rotor-stator method and RNA was extracted (RNeasy micro kit, Qiagen, catalogue number 74004, including the DNase I step). Complementary DNA was then synthesized (High Capacity cDNA kit, Applied Biosystems, catalogue number 4368814) and analysed with an ABI light cycler with the following primers: cyclophilin forward, 5'-TGGAGAGCACCAAGACAGACA-3', and reverse, 5'-TGCCGGAG-TCGACAATGAT-3'; *Fgf8* forward, 5'-TCTCCAGCAGCATCTCTGTGAA-3', and reverse, 5'-GGAAGCTAATTGCCAAGAGCAA-3'; *Shh* forward, 5'-AGCA-GACCGGCTGATGACT-3', and reverse, 5'-AGAGATGGCCAAGGCATT-AA-3'. No-template controls and reverse transcriptase negative controls were included to detect contaminating genomic DNA. In addition, we performed control experiments in which we made cDNA from 0.5, 1, 2 and 4 limb buds to confirm that we could detect small differences in *Fgf8* and *Shh* expression.

Limb buds from embryos stained with LysoTracker Red as previously described<sup>31</sup>, were embedded in 4% low-melt agarose and sectioned (25 μm) on a Leica vibrating microtome. Serial transverse sections were collected, counterstained with 4,6-diamidino-2-phenylindole (DAPI), mounted on slides, and photographed on a Zeiss Axiophot fluorescence microscope or a Nikon Spectral confocal microscope at the UCSF Nikon Imaging Center. For each section, the LysoTracker-Red-positive area was measured by setting a black/white threshold level in the image of the limb bud in that section and counting white pixels in the mesenchyme. The area was measured by counting the total number of pixels in the image of the limb bud. Data from each section were summed across the sample (voxels) to obtain the total amount of cell death and the size (volume) of each limb bud.

31. Grieshammer, U. *et al.* FGF8 is required for cell survival at distinct stages of nephrogenesis and for regulation of gene expression in nascent nephrons. *Development* **132**, 3847–3857 (2005).



## LETTERS

# Free choice activates a decision circuit between frontal and parietal cortex

Bijan Pesaran<sup>1,3</sup>, Matthew J. Nelson<sup>2</sup> & Richard A. Andersen<sup>2,3</sup>

We often face alternatives that we are free to choose between. Planning movements to select an alternative involves several areas in frontal and parietal cortex<sup>1–11</sup> that are anatomically connected into long-range circuits<sup>12</sup>. These areas must coordinate their activity to select a common movement goal, but how neural circuits make decisions remains poorly understood. Here we simultaneously record from the dorsal premotor area (PMd) in frontal cortex and the parietal reach region (PRR) in parietal cortex to investigate neural circuit mechanisms for decision making. We find that correlations in spike and local field potential (LFP) activity between these areas are greater when monkeys are freely making choices than when they are following instructions. We propose that a decision circuit featuring a sub-population of cells in frontal and parietal cortex may exchange information to coordinate activity between these areas. Cells participating in this decision circuit may influence movement choices by providing a common bias to the selection of movement goals.

According to theories of decision making, we make choices by selecting the alternative that is most valuable to us<sup>13</sup>. How much we value each alternative is revealed by our choices. If we value swimming as much as running, we will choose to do both instead of always choosing one over the other. Although actions with similar values can lead to different choices, only one choice can be made at a time. Planning a movement to select an alternative activates many areas of the brain. How does the brain decide what to do? PMd and PRR plan reaching arm movements<sup>14</sup> and are directly connected<sup>12</sup>. We therefore studied these areas to identify a neural circuit for deciding where to reach. We trained two monkeys to do a free search task and an instructed search task (Fig. 1a, b). In both tasks, monkeys made a sequence of reaches to visual targets for rewards of juice. The key manipulation was that, in the free search task, the three targets were visually identical circles, and the monkey could search in any sequence (Fig. 1a); whereas in the instructed search task, the three targets were a circle, a square and a triangle, and the monkey had to search in a fixed sequence (Fig. 1b). To control other sensory, motor and reward-related factors, we carefully matched the two tasks by yoking the sequences presented in the instructed task to the monkey's choices in the free search task (see Methods, Supplementary Results and Supplementary Fig. 2).

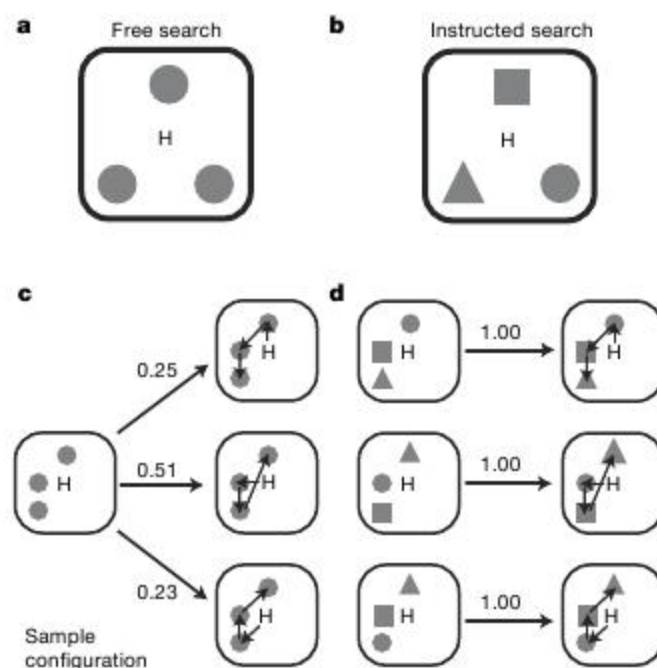
During free search, each monkey's choices varied, even for identical stimuli (Fig. 1c). In contrast, instructed search movement sequences did not vary (Fig. 1d). Overall, each monkey developed a free search strategy and chose between two or three different movement sequences for most search arrays (Supplementary Fig. 3).

Although the tasks we studied could differ in other aspects, like reward expectancy, attention or overall effort, analysis of each animal's behaviour indicates that the major difference involves decision making (Supplementary Results). Free and instructed search involve

different decisions because the alternatives have different values. Free search involves choosing between movement sequences with similar values so choices vary from trial to trial (Fig. 1c). Because we reward only one movement sequence, instructed search involves alternatives with very different values. Consequently, each monkey repeatedly makes the same choices (Fig. 1d).

When movement choices vary from trial to trial, PMd and PRR must coordinate their activity. Analysing spiking and LFP activity may resolve neural coordination between these areas. Spiking activity measures action potentials from individual neurons. LFP activity predominantly measures synaptic potentials in a population of neurons near the recording electrode<sup>15</sup>. Spike-field coherency directly relates these two signals by measuring how well LFP activity is predicted by action potentials. We therefore measured spike-field coherency to characterize neural coordination between PMd and PRR and identify the neurons involved in this coordination.

We made 314 PMd spike–PRR field and 187 PRR spike–PMd field recordings in two animals during both free and instructed search



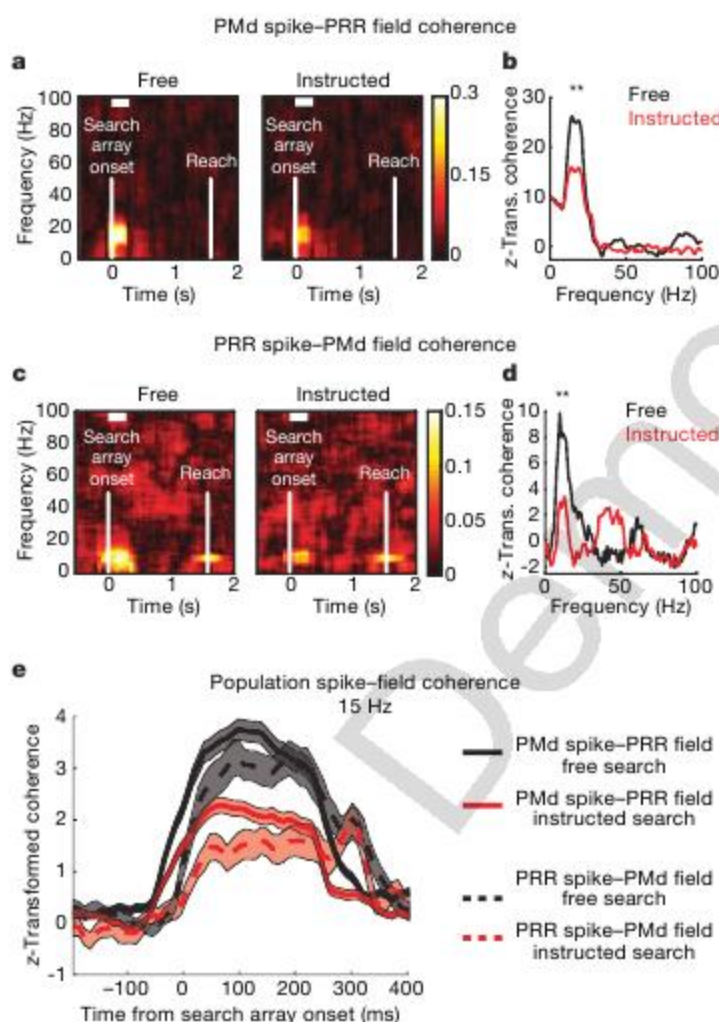
**Figure 1 | Task and behaviour.** **a**, Free search task. Three circular targets presented at eight potential locations spaced 10° apart around the central hand position, H. **b**, Instructed search task. Targets in the instructed search task were a circle, square and triangle; the monkey had to reach to them in that order. Each target had an equal, one-third, probability of being the rewarded target. **c**, The most frequent movement sequences made in response to an example configuration during the free search task. The same configuration elicits three different sequences. **d**, Instructed search configurations elicit the same sequence. Probability is shown above each arrow.

<sup>1</sup>Center for Neural Science, New York University, New York, New York 10003, USA. <sup>2</sup>Computation and Neural Systems Program, California Institute of Technology, Pasadena, California 91125, USA. <sup>3</sup>Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.



tasks (Supplementary Materials). We estimated spike–field coherence between spiking in PMd and LFP activity in PRR using a  $\pm 150$  ms analysis window that was stepped through the trial every 10 ms from before the onset of the search array to the time of the first reach. A highly significant, transient increase in 15-Hz coherence after search array onset was clearly present, as illustrated in an example recording in Fig. 2a. Coherence was significant during both tasks but stronger during free search (Fig. 2b;  $P < 0.05$ ,  $t$ -test). Coherence between spiking in PRR and LFP activity in PMd revealed a similar pattern (Fig. 2c, d). In this recording, coherence was only significant during free search and not during instructed search (Fig. 2d).

Significant coherence at 15 Hz implies that the timing of action potentials is correlated with fluctuations in LFP activity. Analysing the relative phase of activity in PMd and PRR supported this and revealed correlations in the timing of activity in each area that were not simply time-locked to search array onset (see Supplementary Results and Supplementary Fig. 4). Interestingly, the amplitude of spike and LFP activity, as opposed to their relative timing, did not predict PMd–PRR coherence. We correlated the strength of the coherence immediately after search array onset with LFP power and did not observe a significant correlation ( $P = 0.45$ ;  $F$ -test).



**Figure 2 | PMd–PRR spike–field coherence.** **a, b**, Example PMd spike–PRR field coherence: **a**, Time–frequency coherence every 50 ms during free and instructed search. Amplitude is colour coded. Activity is aligned to search array onset (first vertical white bar). Average time of the first reach (second vertical white bar). White horizontal bar shows analysis window for **b**. **b**, Coherence line plot for free (black) and instructed (red) search tasks. Coherence is z-transformed. Significant difference at 15 Hz ( $**P < 0.05$ ;  $t$ -test). **c, d**, Example PRR spike–PMd field coherence. **e**, Population average 15 Hz PMd–PRR spike–field coherence every 10 ms. PMd spike–PRR field coherence (solid); PRR spike–PMd field coherence (dashed). Free search (black); instructed search (red). Coherence is z-transformed before averaging; 95% confidence intervals, Bonferroni-corrected (shaded).

**Table 1 | Population PMd spike–PRR field coherence**

Centre-out	Free or instructed		
23/221 (10%)	74/314 (24%)		
	Free only	Instructed only	Free and instructed
	31/74 (42%)	20/74 (27%)	22/74 (22%)

Linear regression of spike–field coherence against the change in firing rate immediately after search array onset also revealed that coherence was not simply related to the firing rate ( $r^2 = 0.06$ ,  $P = 0.14$ ). Cells with an increase in firing rate generally had the greatest coherence. However, coherence also increased for some cells whose firing rate decreased or did not change.

Spike–field correlations were present only between select pairs of recording sites. Across the population, 74 PMd spike–PRR field recordings (74/314, 24%) contained statistically significant coherence at 15 Hz after search array onset during either task ( $P < 0.05$ ;  $t$ -test; Table 1). A similar proportion of PRR spike–PMd field recordings were significant (43/187, 23%;  $P < 0.05$ ; Table 2). In both cases, spike–field coherence was most prevalent during free search. The fraction of correlated recordings significantly increased between sites with overlapping (less than  $20^\circ$ ) response fields ( $P < 0.05$ ; binomial test; 54% of PMd spike–PRR field recordings, 45% of PRR spike–PMd field recordings).

To test whether spike–field coherence between PMd and PRR is specific to decision making, we measured coherence during two control experiments. First, we measured spike–field coherence during a single-target centre-out task, instructing monkeys to move to a single peripheral target. In this task, there was no choice between targets. The proportions of recordings with significant spike–field coherence fell dramatically (Tables 1 and 2). Only 10% (23/221) of PMd spike–PRR field recordings and 9% (13/138) of PRR spike–PMd field recordings had significant coherence. Second, during both search tasks, we found that saccades are reliably made after search array onset (see Supplementary Results). To test whether spike–field coherence was due to these eye movements, we measured coherence in one animal during a variant of the search tasks that involved enforced fixation. Even during fixation, spike–field coherence was significant after search array onset and was strongest during free search (Supplementary Results, Supplementary Fig. 5 and Supplementary Tables 1 and 2). The population average spike–field coherence across all cells recorded during each task reinforced the selectivity for the search tasks (Supplementary Fig. 6). These control experiments demonstrate that spike–field coherence between PMd and PRR is associated with making a decision.

LFP activity was not only correlated with spiking activity in the other area. Within-area spike–field coherence was also significant (Supplementary Results and Supplementary Figs 7 and 8). Because spiking was coherent with locally recorded LFP activity, correlations in LFP activity between areas may capture the correlation we observe. Partial spike–field coherence analysis<sup>16</sup> addresses this concern (Supplementary Methods and Supplementary Fig. 9). In each example recording, partial spike–field coherence remained significant after accounting for local LFP activity ( $P < 0.05$ ,  $t$ -test). Significant partial spike–field coherence was also present across the population (74% of PMd spike–PRR field partial coherence and 70% of PRR spike–PMd field partial coherence; see Supplementary Results). Therefore, spike–field coherence between PMd and PRR directly relates the activity of individual neurons with distant LFP activity.

**Table 2 | Population PRR spike–PMd field coherence**

Centre-out	Free or instructed		
13/138 (9%)	43/187 (23%)		
	Free only	Instructed only	Free and instructed
	21/43 (49%)	12/43 (28%)	9/43 (21%)



Spike-field coherence gives two independent measures of the neuronal coordination between PRR and PMd. This may indicate how activity flows across the circuit. We estimated the population average coherence for each of the populations that showed coherence at 15 Hz in either search task and compared them (Fig. 2e). Across each population, PMd-PRR spike-field z-score coherence (see Supplementary Methods) was stronger during free search than instructed search ( $P < 0.01$ , Bonferroni-corrected *t*-test). Importantly, PMd spike-PRR field coherence started about 30 ms earlier than PRR spike-PMd field coherence.

Assuming that LFP activity is predominantly synaptic, this suggests that PMd is activated before PRR during search and that PMd spiking is reflected in PRR LFP activity before PRR spiking is reflected in PMd LFP activity (see Supplementary Discussion). The activity is at a relatively low frequency, about 15 Hz, and is transient, about 350 ms. Our time resolution is limited, but the correlation can involve only a few 15-Hz cycles. Because action potentials are propagated between areas, one attractive possibility is that spike-field coherence measures signals in a sub-population of neurons that travel across this circuit first from PMd to PRR and then back from

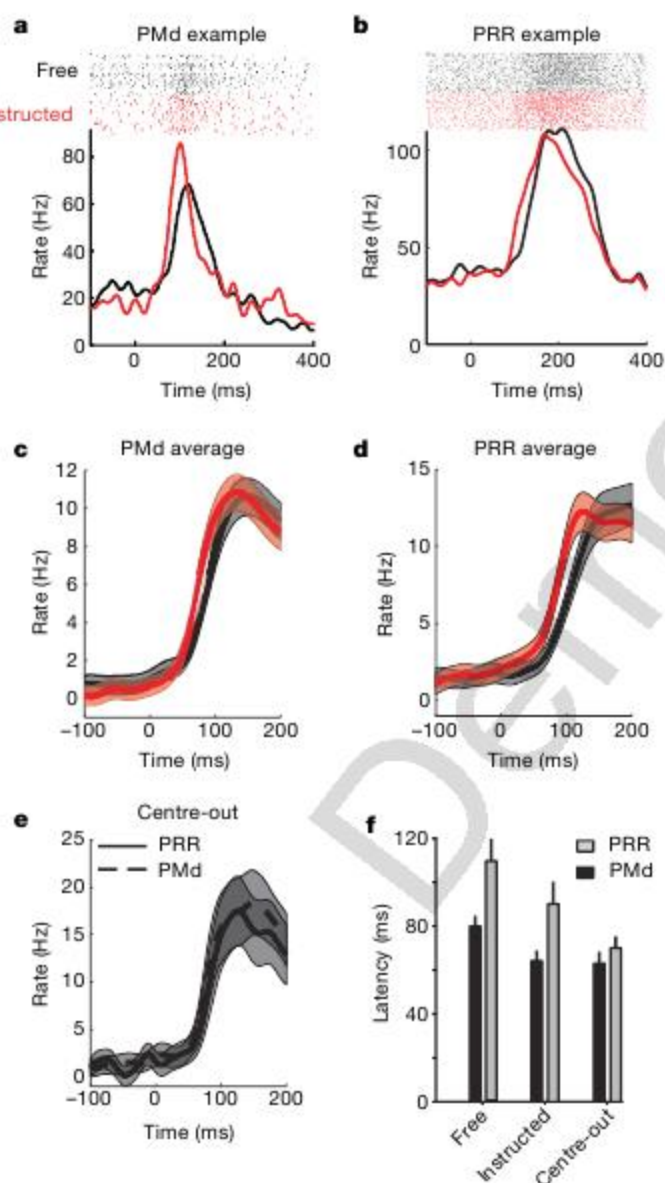
PRR to PMd in a 'handshake'. Consistent with this possibility, the 30 ms latency between the spike-field coherence measurements (Fig. 2e) is a half-cycle at 15 Hz.

PMd and PRR spiking activity lets us examine when each area becomes active. We recorded 115 PMd and 39 PRR neurons responsive to search array onset to measure response latency in each area. PMd spiking responded significantly earlier than PRR spiking in both search tasks (PMd instructed search,  $64 \pm 6$  ms (mean  $\pm$  s.e.m.); free search,  $79 \pm 5$  ms. PRR instructed search,  $90 \pm 10$  ms; free search,  $109 \pm 11$  ms; Fig. 3). We then estimated response latency for 110 PMd neurons and 120 PRR neurons recorded in both animals during the centre-out task. PRR cue response latencies were significantly shorter in this task than in either of the search tasks ( $P < 0.05$ ; permutation test); and PMd and PRR response latencies did not differ (PMd,  $63 \pm 5$  ms; PRR,  $70 \pm 6$  ms;  $P = 0.51$ , Wilcoxon test; Fig. 3e). This suggests that the response latency difference between PRR and PMd is specific to making a decision.

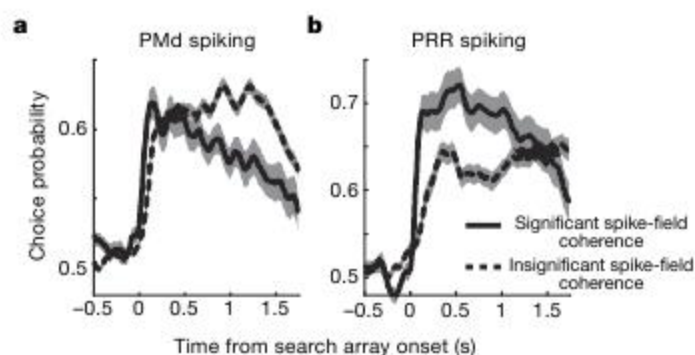
Because spike-field correlations are strongest during decision making, the sub-population of coherent neurons may encode the upcoming movement choice. If so, cells with significant spike-field correlations should predict the movement choice earlier than cells that do not. We analysed this with a receiver-operating characteristic of the firing rate during free search (Supplementary Methods). We calculated the average choice probability separately for correlated and uncorrelated PRR and PMd neurons. In both areas, correlated neurons predict the movement choice after search array onset during the period of greatest spike-field coherence (Fig. 4a, b; see also Fig. 2e). Later in the trial, uncorrelated cells predict the movement choice as accurately as correlated cells. Neurons with long-range correlations may, therefore, exchange information about movement choice between PMd and PRR.

In summary, correlations between PMd and PRR are activated by decision making. Coherence is strongest during free search and is weaker during instructed search. Far less coherence is present during a simpler centre-out task, and the pattern of coherence is unaffected by freely made eye movements. This shows that decision making is distributed across a frontal-parietal circuit and that top-down signals from PMd influence decisions in this circuit.

Why is coherence stronger during free search? This could be due to the nature of the decision. Choices were variable during free search. In contrast, the same choices were made repeatedly during instructed search (Fig. 1). Decision making can be modelled by races underlying the selection of each alternative<sup>17</sup>. These races must be closer during free search because choices are more variable. Therefore, the difficulty of the decision may underlie coherence between PMd and PRR. Cognitive control mechanisms are activated to select between alternative actions. Prefrontal, medial frontal and cingulate cortex are involved in these mechanisms<sup>18–20</sup> and could modulate frontal-parietal coherence during decision making.



**Figure 3 | Spike response latencies.** **a**, Example PMd neuron response to free search (black) and instructed search (red). Activity is aligned to search array onset. Movement to the cell's preferred direction. **b**, Example PRR neuron. **c**, Population average PMd spike response for cells. Activity is baseline subtracted; s.e.m. (shaded). **d**, Population average PRR spike response. **e**, Population average PRR and PMd spike responses during centre-out task to the preferred direction. **f**, Population response latencies for PMd and PRR during free search, instructed search and the centre-out task. Error bars, 95% confidence intervals.



**Figure 4 | Receiver-operating characteristic choice probability estimated from the firing rate for neurons with and without significant PMd-PRR spike-field coherence.** **a**, Population average choice probability for correlated (solid) and uncorrelated (dashed) PMd neurons; 95% confidence intervals (shaded). **b**, Same for PRR neurons.



During search, the flow of activity across frontal and parietal cortex may reflect the process of deciding. Information rises fastest in PMd (Fig. 3f), so it cannot be driven by PRR<sup>14</sup> and must take alternative routes, possibly through the thalamus<sup>21</sup>. Information may go from frontal to parietal cortex<sup>22</sup> and then back in a 'handshake' of increased communication (Fig. 2e) that reflects the decision. This transient coordination may reflect how long the decision takes. Subsequent activity may reflect movement planning after the decision (see Supplementary Discussion). Oscillations and synchronization in frontal and parietal cortex exist during attention and movement preparation<sup>4,23–28</sup>. Correlations at specific frequencies could be a signature of these cognitive processes<sup>29</sup>. We have identified a decision circuit in which frontal–parietal communication occurs at relatively low frequencies. The neurons participating in this circuit could play an important role in deciding where to reach.

## METHODS SUMMARY

Two male rhesus monkeys (*Macaca mulatta*) participated in the experiments. We recorded single-unit and LFP activity from PMd and PRR using Pt/Ir electrodes controlled by multiple-electrode microdrives (Thomas Recordings). Each monkey was trained to perform a reach search for juice rewards either by freely making choices or by following instructions. Correlations between spiking and LFP activity within and between PMd and PRR were estimated using multitaper spectral methods<sup>4,30</sup>. All surgical and animal care procedures were done in accordance with National Institutes of Health guidelines and were approved by the California Institute of Technology Animal Care and Use Committee.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 January 2007; accepted 22 February 2008.  
Published online 16 April 2008.

- Romo, R. & Schultz, W. Neuronal activity preceding self-initiated or externally timed arm movements in area 6 of monkey cortex. *Exp. Brain Res.* **67**, 656–662 (1987).
- Platt, M. L. & Glimcher, P. W. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238 (1999).
- Gold, J. I. & Shadlen, M. N. Representation of a perceptual decision in developing oculomotor commands. *Nature* **404**, 390–394 (2000).
- Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P. & Andersen, R. A. Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nature Neurosci.* **5**, 805–811 (2002).
- Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Matching behavior and the representation of value in the parietal cortex. *Science* **304**, 1782–1787 (2004).
- Cisek, P. & Kalaska, J. F. Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron* **45**, 801–814 (2005).
- Gail, A. & Andersen, R. A. Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. *J. Neurosci.* **26**, 9376–9384 (2006).
- Pesaran, B., Nelson, M. J. & Andersen, R. A. Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning. *Neuron* **51**, 125–134 (2006).
- Quiroga, R., Snyder, L. H., Batista, A. P., Cui, H. & Andersen, R. A. Movement intention is better predicted than attention in the posterior parietal cortex. *J. Neurosci.* **26**, 3615–3620 (2006).
- Scherberger, H. & Andersen, R. A. Target selection signals for arm reaching in the posterior parietal cortex. *J. Neurosci.* **27**, 2001–2012 (2007).
- Yang, T. & Shadlen, M. N. Probabilistic reasoning by neurons. *Nature* **447**, 1075–1080 (2007).
- Johnson, P. B., Ferraina, S., Bianchi, L. & Caminiti, R. Cortical networks for visual reaching: physiological and anatomical organization of frontal and parietal lobe arm regions. *Cereb. Cortex* **6**, 102–119 (1996).
- Kreps, D. M. *A Course in Microeconomic Theory* Ch. 2 (Princeton Univ. Press, Princeton, 1990).
- Wise, S. P., Boussaoud, D., Johnson, P. B. & Caminiti, R. Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annu. Rev. Neurosci.* **20**, 25–42 (1997).
- Mitzdorf, U. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* **65**, 37–100 (1985).
- Halliday, D. M. et al. A framework for the analysis of mixed time series/point process data—theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Prog. Biophys. Mol. Biol.* **64**, 237–278 (1995).
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
- Goldberg, G. Supplementary motor area structure and function: review and hypotheses. *Behav. Brain Sci.* **8**, 567–616 (1985).
- Kerns, J. G. et al. Anterior cingulate conflict monitoring and adjustments in control. *Science* **303**, 1023–1026 (2004).
- Daw, N. D. & Doya, K. The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* **16**, 199–204 (2006).
- Schmolesky, M. T. et al. Signal timing across the macaque visual system. *J. Neurophysiol.* **79**, 3272–3278 (1998).
- Cisek, P. Integrated neural processes for defining potential actions and deciding between them: a computational model. *J. Neurosci.* **26**, 9761–9770 (2006).
- Bressler, S. L., Coppola, R. & Nakamura, R. Episodic multiregional cortical coherence at multiple frequencies during visual task-performance. *Nature* **366**, 153–156 (1993).
- Murthy, V. N. & Fetz, E. E. Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys. *J. Neurophysiol.* **76**, 3968–3982 (1996).
- Riehle, A., Grun, S., Diesmann, M. & Aertsen, A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 1950–1953 (1997).
- Scherberger, H., Jarvis, M. J. R. & Andersen, R. A. Cortical local field potential encodes movement intentions. *Neuron* **46**, 347–354 (2005).
- Rickert, J. et al. Encoding of movement direction in different frequency ranges of motor cortical local field potentials. *J. Neurosci.* **25**, 8815–8824 (2005).
- Buschman, T. J. & Miller, E. K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* **315**, 1860–1862 (2007).
- Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9**, 474–480 (2005).
- Mitra, P. P. & Pesaran, B. Analysis of dynamic brain imaging data. *Biophys. J.* **76**, 691–708 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the National Eye Institute, the National Institute of Mental Health, the Defense Advanced Research Projects Agency BioInfoMicro program, a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund (B.P.), a James D. Watson Investigator Program Award from NYSTAR (B.P.) and a Sloan Research Fellowship (B.P.). We thank: N. Daw, H. Dean and D. Heeger for comments; T. Yao for editorial assistance; K. Pejsa and N. Sammons for animal care; and V. Shcherbatyuk and M. Walsh for technical assistance.

**Author Contributions** B.P., M.J.N. and R.A.A. designed the experiment and wrote the paper. B.P. and M.J.N. collected the data. B.P. performed the data analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.P. ([bijan@nyu.edu](mailto:bijan@nyu.edu)).



## METHODS

**Experimental preparation.** Two male rhesus monkeys (*Macaca mulatta*) participated in the experiments. Each animal was first implanted with a head cap and eye coil under general anaesthesia. In a second surgery, recording chambers were implanted in frontal and posterior parietal cortex in the right hemisphere of each animal. Structural magnetic resonance imaging identified the position of the arcuate sulcus and intraparietal sulcus and guided placement of the recording chambers to give access to cortex medial to each sulcus. In both animals, PMd recordings were made within the cortical gyrus within 1.5 mm of the cortical surface, and PRR recordings were made within the intraparietal sulcus 4–9 mm below the cortical surface.

**Behavioural tasks.** For all tasks, reaches were made with the left arm on a touch-sensitive screen (ELO Touch Systems). Visual stimuli were presented on an LCD display (LG Electronics) placed behind the touch screen. All trials began with the illumination of a central circle which the animal needed to touch with his hand and hold for a baseline period (about 500 ms).

In the search tasks, after a baseline hold period (0.5–1 s), three targets were presented on a  $3 \times 3$  grid (spaced  $10^\circ$ ) of eight possible locations around the start point. After a delay period (1–1.5 s) the monkey was given a 'go' signal to reach to one of the three targets. Only one of the three targets triggered a juice reward when touched. If the monkey did not reach to the target that gave the reward, he was allowed to make additional reaches to targets after subsequent hold periods (0.5–1 s). Additional reaches were allowed until the reward was received. Targets were extinguished once they were touched. An auditory tone signalled the 'go' signal for each reach. A different set of three targets from the eight possible locations appeared for each trial, and the target that gave the reward was chosen from these three targets with equal probability. This stimulus–reward configuration set ensured that the monkey did not repeatedly perform the same stereotyped sequence of movements. This elicited choices by releasing constraints instead of intensively training the subject to overcome biases and avoid stereotyped choices. If the animal reached for the wrong shape in the instructed search task, the trial was aborted. The animal first knew it was in a free search or instructed search trial when the search array was illuminated.

The free and instructed search tasks were yoked in an interleaved design to match the sensory-, motor- and reward-related contingencies. We did this by requiring the monkey to perform an initial set of free search trials in a block (typically 50). The search array configurations were selected at random from the set of 56 possible configurations. We counted the number of times each search array configuration was presented and the number of times each possible movement sequence was made during the free search task. After the initial set of free search trials was performed, we began to randomly interleave instructed search trials. During this phase of the session, the probability of a given trial being a free search or instructed search task was balanced so that after 200 total trials an equal number of trials from each task would be successfully completed. Search array configurations for the free search task continued to be selected at random. Search array configurations for the instructed search task were drawn from the probability distribution defined by the set of search configurations presented in the preceding free search trials that were successfully completed. To match the motor contingencies in the instructed search trials to the free search trials, the order of the movement sequences instructed by the search array was drawn from a probability distribution defined by the set of movement choices made in the preceding free search trials. To reduce the number of trials needed to estimate these movement sequence probabilities and to prevent the generation of stereotyped movement sequences, we matched only the first element of the instructed movement sequence with the monkey's choices and allowed potential mismatch for the second and third elements of the instructed movement sequence. All probability distributions were updated after each successful trial. Eye movements were unconstrained and, on a subset of experimental sessions (53 sessions in monkey E, 15 sessions in monkey Z), were monitored using a scleral search coil (CNC Engineering).

A variant of the search tasks with enforced fixation was also tested in one animal (monkey E). In this variant, the search tasks were identical except that the monkey needed to maintain fixation at the current touch location throughout the trial. As a result, the only eye movements that were allowed were made at the time of a reach movement.

In the centre-out task, a single target was presented at one of eight peripheral locations on a  $3 \times 3$  grid (spaced  $10^\circ$ ) around the start point. After a delay period (1–1.5 s) the monkey reached for the target and was then given a juice reward. Fixation was enforced during the period after acquisition of the start point until the end of the delay period. At this time, gaze was unconstrained and both monkeys made a coordinated saccade to the target of the reach movement.



## LETTERS

# Vascular normalization in *Rgs5*-deficient tumours promotes immune destruction

Juliana Hamzah<sup>1</sup>, Manfred Jugold<sup>2</sup>, Fabian Kiessling<sup>2</sup>, Paul Rigby<sup>5</sup>, Mitali Manzur<sup>1</sup>, Hugo H. Marti<sup>6</sup>, Tamer Rabie<sup>6</sup>, Sylvia Kaden<sup>3</sup>, Hermann-Josef Gröne<sup>3</sup>, Günter J. Hämmerling<sup>4</sup>, Bernd Arnold<sup>4</sup> & Ruth Ganss<sup>1</sup>

The vasculature of solid tumours is morphologically aberrant and characterized by dilated and fragile vessels, intensive vessel sprouting and loss of hierarchical architecture<sup>1</sup>. Constant vessel remodelling leads to spontaneous haemorrhages<sup>2</sup> and increased interstitial fluid pressure in the tumour environment<sup>3,4</sup>. Tumour-related angiogenesis supports tumour growth and is also a major obstacle for successful immune therapy as it prevents migration of immune effector cells into established tumour parenchyma<sup>2,5,6</sup>. The molecular mechanisms for these angiogenic alterations are largely unknown. Here we identify regulator of G-protein signalling 5 (*Rgs5*) as a master gene responsible for the abnormal tumour vascular morphology in mice. Loss of *Rgs5* results in pericyte maturation, vascular normalization and consequent marked reductions in tumour hypoxia and vessel leakiness. These vascular and intratumoral changes enhance influx of immune effector cells into tumour parenchyma and markedly prolong survival of tumour-bearing mice. This is the first demonstration, to our knowledge, of reduced tumour angiogenesis and improved immune therapeutic outcome on loss of a vascular gene function and establishes a previously unrecognized role of G-protein signalling in tumour angiogenesis.

During tumour-induced angiogenesis, the two important vascular cell types, endothelial cells and surrounding pericytes, develop multiple morphological and architectural abnormalities<sup>7,8</sup> as well as altered expression of marker proteins<sup>9–12</sup>. Using the RIP-Tag mouse model of pancreatic islet carcinogenesis (SV40 large T antigen expressed under the control of the rat insulin gene (*Ins2*) promoter), we have recently observed that *Rgs5* is overexpressed in the aberrant tumour vasculature<sup>10</sup>.

*Rgs5* is expressed by pericytes in the vascular bed and is the first marker for a subgroup of platelet-derived growth factor receptor  $\beta$  (PDGFR $\beta$ )<sup>+</sup> progenitor perivascular cells that regulate vascular survival in tumours<sup>8</sup>. RGS molecules are a family of biochemically well-characterized molecules that inhibit signalling from G-protein-coupled receptors by stimulating the intrinsic GTPase activity of activated G $\alpha$  proteins<sup>13</sup>. However, their function *in vivo* is largely unknown. *Rgs5* is constitutively expressed in a variety of organs, especially brain, heart, aorta, skeletal muscle, liver and kidney, and is also upregulated in RIP1-Tag5 tumour vessels<sup>10,14</sup>.

To study the role of RGS5 in tumour angiogenesis and vascular normalization, we generated *Rgs5*-deficient mice by crossing recombinant mice harbouring a loxP-flanked *Rgs5* exon 1 with Cre-deleter mice. On exon 1 deletion, *Rgs5* expression is absent in heart, kidney, brain, lung and liver (Fig. 1a, b). *Rgs5*-deficient mice develop normally and present with no gross histological abnormalities. *Rgs5*<sup>-/-</sup> mice were intercrossed with transgenic RIP1-Tag5 mice to assess

survival and intratumoral characteristics (Supplementary Fig. 1). RIP1-Tag5 mice develop tumours in a well-characterized sequence of events from normal Tag<sup>+</sup> islets to hyperplastic and angiogenic islets, and eventually form insulinomas, which cause premature death due to hypoglycaemia. Early tumorigenesis is unchanged in the *Rgs5*<sup>-/-</sup> background, as documented by a comparable number of angiogenic islets per mouse (Fig. 1c). At later tumour stages, the *Rgs5*-deficient background enhances tumour growth, as shown by an increase in overall tumour burden over age and premature death caused by hypoglycaemia (Fig. 1d). This is in agreement with the reduced survival of *Rgs5*-deficient RIP1-Tag5 mice compared to RIP1-Tag5 wild-type mice ( $P < 0.0001$ ; range,  $27 \pm 3$  weeks for RIP1-Tag5  $\times$  *Rgs5*<sup>-/-</sup>,  $30 \pm 3$  weeks for RIP1-Tag5; Fig. 1e).

To assess tumour vascular morphology, insulinomas in RIP1-Tag5 wild-type or *Rgs5*-deficient RIP1-Tag5 mice at 27 weeks of age were visualized after lectin perfusion using confocal microscopy (Fig. 2a–c). Notably, the vascular network in *Rgs5*<sup>-/-</sup> insulinomas resembles normal vessels with regard to vessel diameters and distribution (for overview, see Supplementary Fig. 2). This is in contrast with insulinomas in wild-type mice, which display a chaotic vascular architecture with large vessels adjacent to small vessels and a heterogeneous vessel density. This finding demonstrates that loss of *Rgs5* expression results in vessel normalization.

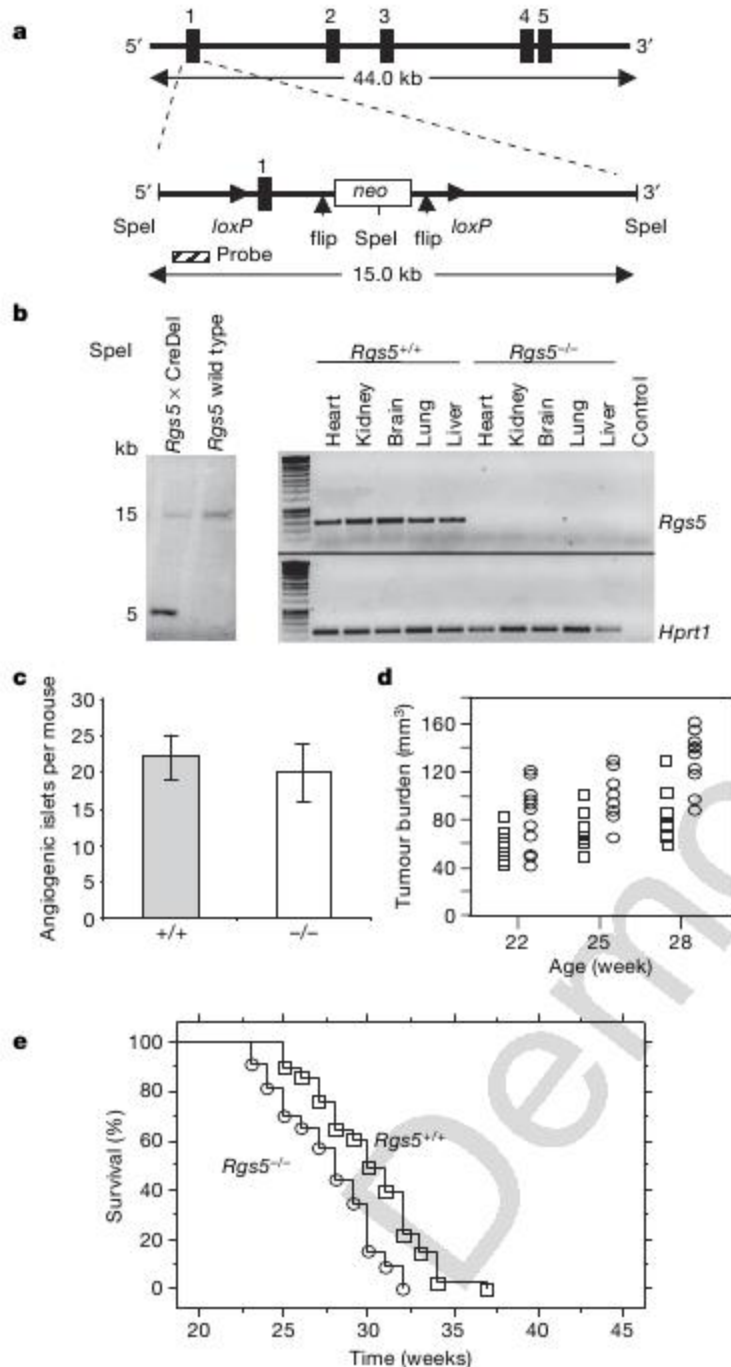
Because vascular *Rgs5* expression is restricted to pericytes, we hypothesized that pericyte maturation, abundance and/or attachment along the vessel wall is intimately associated with the observed vascular normalization. Pericyte coverage and association with endothelial cells were unchanged in *Rgs5*-deficient tumours (Supplementary Fig. 3). However, pericyte phenotypes differed between wild-type and *Rgs5*-deficient tumours. In RIP-Tag wild-type tumours, most pericytes are positive for both PDGFR $\beta$  and *Rgs5*, representing immature progenitor cells with the potential to differentiate into mature pericytes *in vitro*. A smaller subpopulation of mature pericytes that are immunoreactive for NG2 (also known as *Cspg4*), desmin and  $\alpha$ SMA (also known as *Acta2*) is also found in these tumours<sup>8</sup>. By contrast, *Rgs5*-deficient tumour pericytes in age- and size-matched tumour samples predominantly express  $\alpha$ SMA and also NG2—recognized markers of maturity<sup>8</sup> (Fig. 2d, e). These data indicate that in the absence of *Rgs5*, tumour pericytes are of a more mature phenotype.

Diverse functions have been described for pericytes, ranging from haemodynamic regulation to vessel stability and permeability<sup>15</sup>. Because *Rgs5*-deficient mice display notably improved vascular integrity and maturity, we reasoned that concomitant changes in the tumour microenvironment were also likely. Interestingly, *Rgs5* expression has been shown to be increased under hydrostatic

<sup>1</sup>Western Australian Institute for Medical Research, UWA Centre for Medical Research, Perth, Western Australia 6000, Australia. <sup>2</sup>Juniorgroup Molecular Imaging, Department of Medical Physics in Radiology, <sup>3</sup>Department of Cellular and Molecular Pathology, and <sup>4</sup>Department of Molecular Immunology, German Cancer Research Center, 69120 Heidelberg, Germany. <sup>5</sup>Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, Perth, Western Australia 6000, Australia. <sup>6</sup>Institute of Physiology and Pathophysiology, University of Heidelberg, 69120 Heidelberg, Germany.

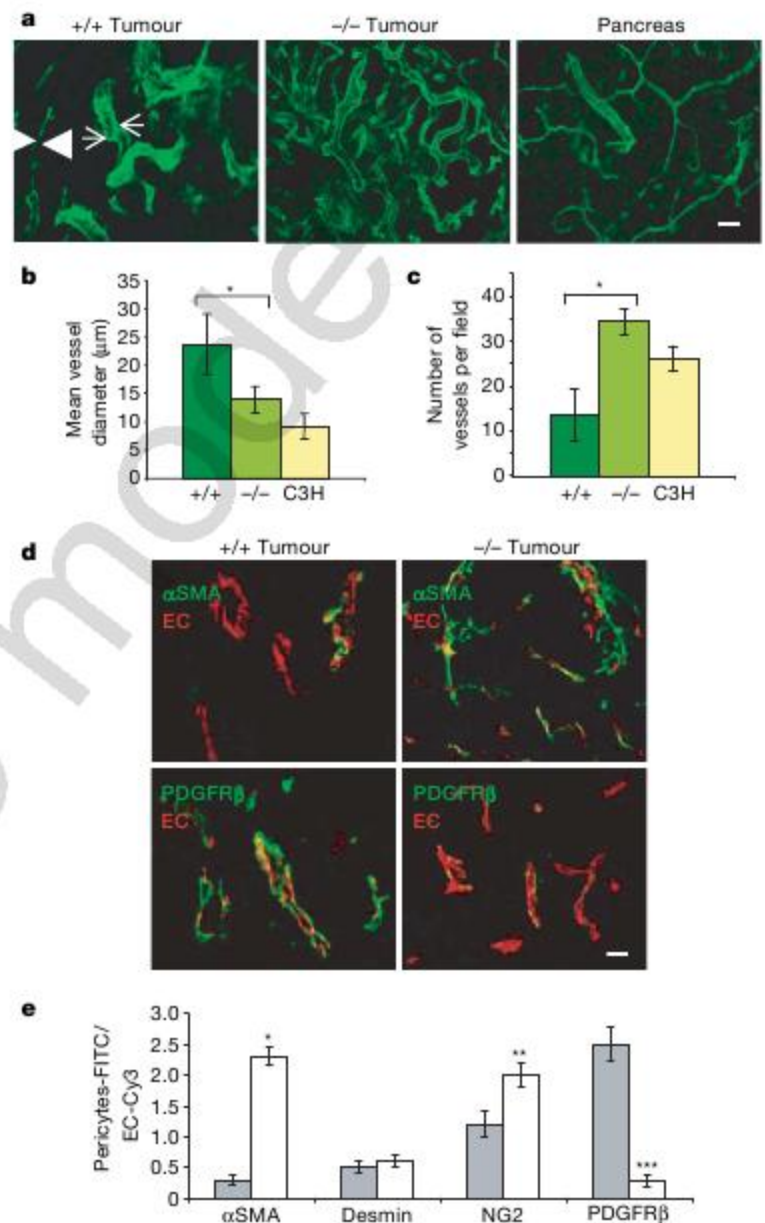


pressure *in vitro*, underscoring a direct correlation between angiogenesis, intratumoural hydrostatic pressure and *Rgs5* expression<sup>16</sup>. We found oxygen supply to be improved in *Rgs5*-deficient tumours compared to wild-type tumours, as demonstrated by a reduced tumour hypoxia; this was visualized by the formation of pimonidazole adducts (Fig. 3a). This finding may contribute to a growth advantage for *Rgs5*-deficient tumours, resulting in poorer survival (Fig. 1e).



**Figure 1 | Establishment of *Rgs5*<sup>-/-</sup> mice.** **a**, *Rgs5* exon 1 was flanked with *loxP* sites, and a neomycin (*neo*) cassette was introduced into intron 1 and flanked by flip sites for removal after the selection process. SpeI restriction sites are shown. **b**, Left: Southern blot analysis of SpeI-digested tail DNA derived from F<sub>1</sub> offspring after crossbreeding with Cre-deleter (CreDel) mice and wild-type controls. Right: PCR with reverse transcription (RT-PCR) analysis of organs from a *Rgs5*<sup>+/+</sup> and a *Rgs5*<sup>-/-</sup> mouse. *Hprt1*, hypoxanthine phosphoribosyltransferase. **c**, The number of angiogenic islets in 18-week-old wild-type (grey bar) and *Rgs5*-deficient (white bar) RIP1-Tag5 mice (*n* = 10). Error bars represent s.e.m. **d**, Tumour burden in wild-type (open squares) and *Rgs5*-deficient (open circles) mice at the age of 22, 25 and 28 weeks (*n* = 10; range: *Rgs5*<sup>+/+</sup>, 41–87 mm<sup>3</sup>, 53–108 mm<sup>3</sup>, 57–132 mm<sup>3</sup>, respectively; *Rgs5*<sup>-/-</sup>, 35–128 mm<sup>3</sup>, 63–135 mm<sup>3</sup> and 82–166 mm<sup>3</sup>, respectively). **e**, Kaplan-Meier survival analysis of RIP1-Tag5 wild-type (+/+), open squares, *n* = 82 and RIP1-Tag5 × *Rgs5*<sup>-/-</sup> mice (open circles, *n* = 82, *P* < 0.0001).

To evaluate further the physiological impact of vascular maturation, we compared vascular permeability of wild-type and *Rgs5*<sup>-/-</sup> tumours using dynamic contrast-enhanced magnetic resonance imaging (MRI) on 27-week-old mice. Notably, normalized tumour vessels in *Rgs5*<sup>-/-</sup> mice showed a 50% reduction in permeability for the contrast agent compared to wild-type RIP1-Tag5 tumours (Fig. 3b, *P* = 0.008). These are the first functional data demonstrating reduced leakiness of angiogenic vessels on loss of gene function *in vivo*. Thus, absence of *Rgs5* results in a reduced number of



**Figure 2 | Vascular normalization in *Rgs5*-deficient tumours.** **a**, Confocal images of lectin-perfused vessels in RIP1-Tag5 wild-type (+/+, *n* = 12) and *Rgs5*-deficient (-/-, *n* = 15) size-matched tumours (40–65 mm<sup>3</sup>), and C3H controls (pancreas; *n* = 10). Arrowheads point at a vessel with small caliber next to a vessel with large caliber (arrows). For these images, a ×20 objective was used; scale bar, 50 μm. **b**, Mean vessel diameters were quantified from wild-type (+/+) or *Rgs5*-deficient (-/-), size-matched tumours and C3H pancreatic tissue (C3H; 5 fields per tumour, 15 tumours, \**P* = 0.0009). **c**, Vessel density within randomly selected fields excluding the tumour periphery (5 fields per tumour, 15 tumours, \**P* < 0.0001). **d**, Confocal images from wild-type (+/+) and *Rgs5*-deficient (-/-) mice, labelled with anti-αSMA (green, upper panels) and anti-PDGFRβ (green, lower panels). Endothelial cells (EC) were stained with anti-CD31 (red). For these images, a ×60 objective was used; scale bar, 20 μm. **e**, Quantitative increase of the number of αSMA-, desmin-, NG2- and PDGFRβ-positive pericytes in wild-type (+/+, grey bars) and *Rgs5*-deficient (-/-, white bars) mice. The ratio of total area of green staining (pericyte markers) to red staining (EC) is provided (5 fields per tumour, 15 tumours, \**P* = 0.0001, \*\**P* = 0.002, \*\*\**P* = 0.0001). All error bars represent s.e.m.



enlarged tumour vessels, increased oxygen supply to the tumour parenchyma and decreased tumour vascular permeability. These key findings were reproduced in a fibrosarcoma transplantation model (AG104A) that was subcutaneously grown in *Rgs5*-deficient mice (Supplementary Fig. 4). *Rgs5* is also a marker for developing pericytes in the brain<sup>17</sup>, and its expression persists in adult brain vessels. We therefore investigated the role of RGS5 in the barrier function of brain capillaries during ischaemia. Interestingly, brain oedema after transient cerebral ischaemia was significantly reduced in *Rgs5*-deficient mice, demonstrating decreased permeability for plasma molecules in brain capillaries (Supplementary Table 1), and strongly supports our intratumoural findings (Fig. 3b).

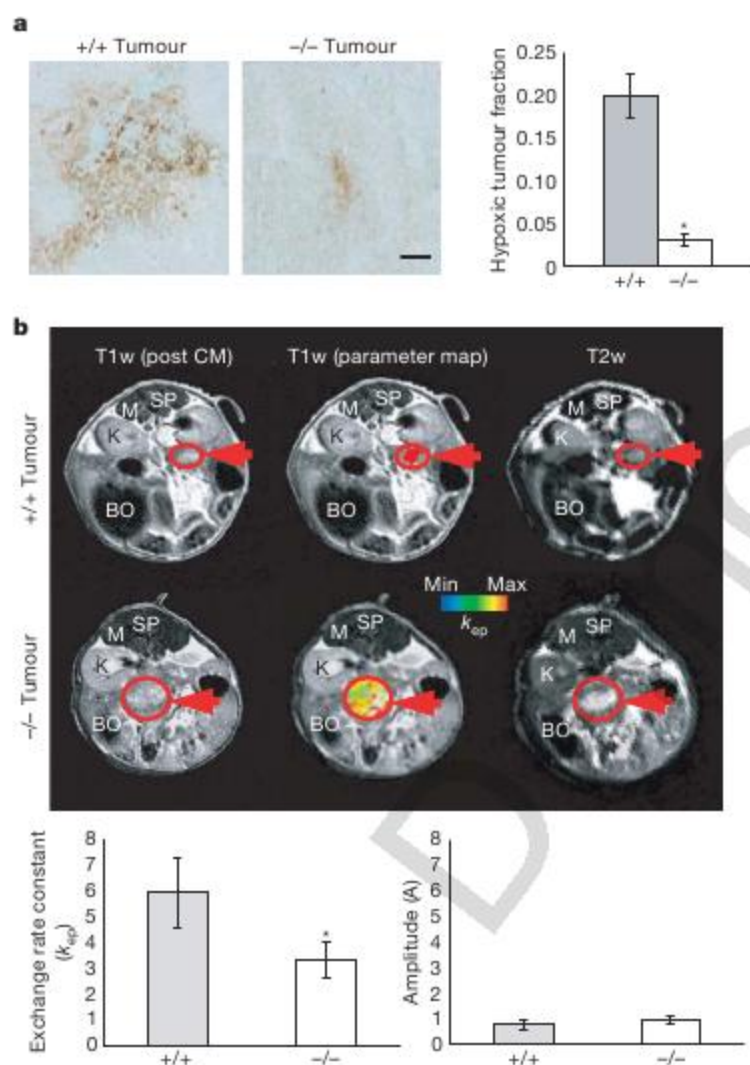
Our previous work on RIP1-Tag5 tumour immunity showed a strong correlation between vascular remodelling induced by intratumoural inflammation and lymphocyte influx<sup>3,18,19</sup>. However, we did not know whether lymphocyte migration into tumour parenchyma was enhanced by vessel remodelling or by an ongoing local immune response. We sought to address this question by adoptively transferring *ex vivo* activated H-2<sup>k</sup>-restricted anti-Tag CD4<sup>+</sup> and CD8<sup>+</sup> T

cells into 27-week-old tumour-bearing RIP1-Tag5  $\times$  *Rgs5*<sup>-/-</sup> mice bred into the C3H background (H-2<sup>k</sup>). Seven days after transfer (the peak of pre-activated Tag-specific T cell proliferation *in vivo*<sup>19</sup>), tumours were harvested and analysed. Untreated tumours in wild-type RIP1-Tag5 mice did not show a spontaneous T-cell infiltrate; by contrast, *Rgs5*-deficient mice consistently displayed a higher degree of spontaneous T-cell infiltration, although this did not reach statistical significance (Fig. 4a and Supplementary Fig. 5). Vessel wall inflammation or upregulation of cytokines and chemokines in untreated *Rgs5*<sup>-/-</sup> tumours were not observed. However, after adoptive transfer (total of  $5 \times 10^6$  CD4<sup>+</sup> and CD8<sup>+</sup> T cells), tumours in the *Rgs5*-deficient mice were massively infiltrated by CD8<sup>+</sup> and CD4<sup>+</sup> T cells, whereas wild-type RIP1-Tag5 tumours showed no significant increase in infiltrating lymphocytes (Fig. 4a and Supplementary Fig. 5). Thus, *Rgs5* loss in angiogenic blood vessels 'opens' tumours for immune cell penetration—a very notable finding further substantiated by impressive survival in subsequent studies.

Cohorts of wild-type and *Rgs5*-deficient RIP1-Tag5 mice (23 weeks old with significant tumour burdens; average range, 50–120 mm<sup>3</sup>) were adoptively transferred with *in vitro* activated CD4<sup>+</sup> and CD8<sup>+</sup> anti-Tag T cells in two-week intervals. In parallel, equal numbers of non-specifically (concanavalin A, ConA) activated or naive T cells from C3H control mice were transferred. As previously reported, adoptive transfer of activated, Tag-specific, naive or ConA-activated lymphocytes alone does not confer a survival advantage to RIP1-Tag5 mice<sup>19,20</sup>, with the mice succumbing to insulinomas at  $32 \pm 2$ ,  $29 \pm 2$  and  $29 \pm 3$  weeks of age, respectively (Fig. 4b–d). However, RIP1-Tag5  $\times$  *Rgs5*<sup>-/-</sup> recipients showed substantially prolonged survival after the transfer of pre-activated, Tag-specific T cells (age  $41 \pm 7$  weeks,  $P = 0.0004$ ; the experiment was terminated when surviving mice reached 48 weeks), correlating with the marked parenchymal influx of anti-Tag immune cells (Fig. 4a), although repeated T-cell infusions were required owing to their limited life-span in a non-inflammatory tumour environment. This survival advantage is tumour-antigen-specific and requires activated anti-Tag T cells because no such effect was seen with transfer of ConA-activated or naive lymphocytes (survival:  $28 \pm 3$  and  $27 \pm 3$  weeks, respectively, Fig. 4c, d). Similarly, *Rgs5*-deficient mice are highly responsive to therapeutic anti-Tag vaccination (survival:  $47 \pm 7$  weeks,  $P < 0.0001$ , started at 23 weeks); in contrast, vaccination strategies in wild-type RIP1-Tag5 mice are only successful in an early prophylactic setting before tumour development (starting at week 6, ref. 20), and fail later, when highly vascularized tumours are established (Fig. 4e and Supplementary Fig. 6).

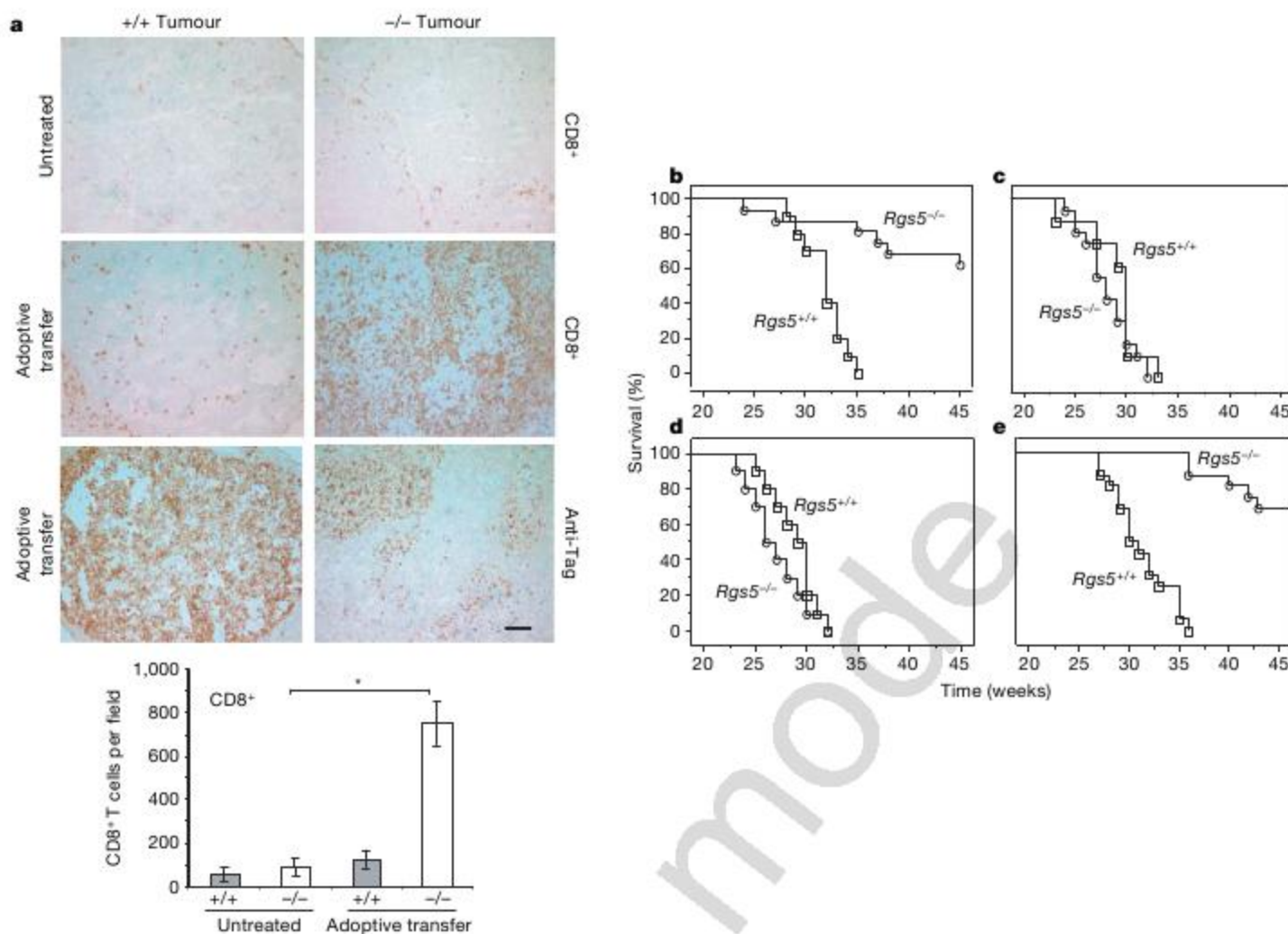
The present study identifies RGS5 as a key regulator controlling the aberrant morphology of the tumour vasculature. Moreover, our results demonstrate the highly dynamic and reversible nature of tumour angiogenesis. Loss of *Rgs5* gene function induces notable morphological and physiological changes in the tumour vasculature and microenvironment. Importantly, deficiency of *Rgs5* reduces angiogenic activity and notably improves the outcome of specific therapeutic interventions.

This study has some overlapping features with anti-angiogenesis therapy in relation to vessel diameters and alleviation of hypoxia. Blocking vascular endothelial growth factor (VEGF) signalling in tumours has been shown to create a 'vascular normalization window' that decreases interstitial pressure and enhances tumour oxygenation and the therapeutic response to cytotoxic drugs and/or radiation in mouse and human cancers<sup>21,22</sup>. *Rgs5* deficiency is, however, clearly distinct with regard to vessel density and results in changes to intratumoural pericyte phenotype rather than pericyte coverage<sup>23</sup>. Furthermore, we show for the first time, to our knowledge, an association between vessel normalization in the absence of *Rgs5* and an increased anti-tumour immune response. Whereas VEGF-blocking therapies clearly enhance drug efficacy, correlating with changes in vascular morphology, pericyte maturation shown in this study may



**Figure 3 | Improved oxygenation and reduced vessel leakiness in *Rgs5*<sup>-/-</sup> tumours.** **a**, Hypoxia in 27-week-old RIP1-Tag5 (+/+) and RIP1-Tag5  $\times$  *Rgs5*<sup>-/-</sup> (-/-) mice ( $n = 20$  tumours, size matched, range of 40–65 mm<sup>3</sup>,  $P < 0.0001$ ). For these images, a  $\times 20$  objective was used; scale bar, 50  $\mu$ m. Error bars represent s.e.m. **b**, Photographs represent T1-weighted (T1w; post CM, post contrast media) MRI images and the corresponding colour-coded parameter maps of the exchange rate constant  $k_{ep}$  (indicating vessel permeability), and T2-weighted MRI images of RIP1-Tag5 wild-type (+/+, upper panel) and RIP1-Tag5  $\times$  *Rgs5*<sup>-/-</sup> (-/-, lower panel) mice recorded with a 1.5 tesla (T) whole-body MR scanner in combination with a small animal coil. Arrows indicate location of tumours. BO, bowel; K, kidney; M, dorsal muscle; SP, spine. Graphs represent quantitative analyses of MRI studies (one-sided *t*-test):  $k_{ep}$ , exchange rate constant; A, amplitude indicating relative blood volume ( $n = 10$ , \* $P = 0.008$ ). Error bars represent s.d.





**Figure 4 | Immune-mediated tumour rejection after vascular normalization.** **a**, Spontaneous infiltration of CD8<sup>+</sup> immune cells measured in tumours of 27-week-old RIP1-Tag5 wild-type (+/+) and *Rgs5*-deficient (-/-) mice (upper images). Infiltration of CD8<sup>+</sup> T cells after adoptive transfer into 27-week-old RIP1-Tag5 wild-type and *Rgs5*<sup>-/-</sup> mice (middle images) and corresponding anti-Tag tumour staining (lower images). For these images, a  $\times 10$  objective was used; scale bar, 100  $\mu$ m. The graph shows quantification of CD8<sup>+</sup> lymphocytic infiltration in size-matched RIP1-Tag5 wild-type (grey bars) and *Rgs5*-deficient (white bars) tumours without treatment and after adoptive transfer (5 fields per tumour,  $n = 15$  tumours,

$*P < 0.0001$ ). Error bars represent s.e.m. **b-d**, Kaplan-Meier-survival studies on RIP1-Tag5 wild-type (*Rgs5*<sup>+/+</sup>, open squares,  $n = 10-16$ ) and RIP1-Tag5  $\times$  *Rgs5*<sup>-/-</sup> mice (*Rgs5*<sup>-/-</sup>, open circles,  $n = 12-16$ ) treated with: **b**, pre-activated anti-Tag T cells (logrank test,  $P = 0.0004$ ); **c**, non-specific ConA-activated T cells ( $P = 0.27$ ); or **d**, naive T cells from C3H mice ( $P = 0.24$ ). Treatment start, 23 weeks; untreated controls are shown in Fig. 1e. **e**, Survival of RIP1-Tag5 wild-type (open square,  $n = 16$ ) and *Rgs5*-deficient (open circle,  $n = 16$ ) mice after vaccination with Tag and CpG-ODN 1668,  $P < 0.0001$ . Treatment start: 23 weeks.

also contribute to the anti-tumour effects by directly influencing leukocyte attachment and transmigration into tumour parenchyma. There is emerging evidence that resistance develops to anti-VEGF/VEGF receptor therapies, and that re-growth of treated tumours may occur due to compensation by other pro-angiogenic factors<sup>24</sup>. Here we provide novel mechanistic insights into reversing tumour angiogenesis independently of anti-angiogenic drugs by targeting G-protein signalling. Studies are under way to elucidate pathways regulated by RGS5 in tumour pericytes for pharmacotherapeutic intervention in combination with immune therapy. Recognition that RGS5 is a broadly expressed tumour antigen<sup>25</sup> and is also vessel-associated in numerous tumours (for example, astrocytomas and insulinomas<sup>10</sup>, renal cell carcinoma<sup>26</sup> and hepatocellular carcinoma<sup>27</sup>) confirms its general importance in tumorigenesis and expands potential therapeutic opportunities.

## METHODS SUMMARY

*Rgs5*-deficient mice were established by removing exon 1. *Rgs5*<sup>-/-</sup> mice were crossbred with RIP1-Tag5, SV40 large T antigen transgenic mice on a C3H background. Angiogenic islets and tumours were isolated from pancreatic tissue. Confocal microscopy was performed on lectin-perfused, albumin/gelatin-embedded, vibratome-dissected tissue to assess vascular density and vessel diameters. Anti- $\alpha$ SMA, anti-PDGFR $\beta$ , anti-NG2 and anti-desmin histology was used to assess pericyte maturation. Hypoxia in islet tumours was detected by

the formation of pimonidazole adducts after the injection of pimonidazole hydrochloride compound into tumour-bearing mice. MRI was performed using a 1.5 tesla whole-body MR-scanner (Siemens Symphony) in combination with a custom-made radio-frequency coil for excitation and signal reception to assay vascular leakiness. Anti-Tag, CD4<sup>+</sup> and CD8<sup>+</sup> transgenic T cells were used for adoptive transfers to evaluate lymphocyte access into tumours. Vaccination studies were performed using 50  $\mu$ g of Tag protein mixed with 50  $\mu$ g of CpG-oligodeoxynucleotide (ODN) 1668.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 14 December 2007; accepted 25 February 2008.

Published online 16 April 2008.

1. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353-364 (1996).
2. Ryschich, E., Schmidt, J., Hammerling, G. J., Klar, E. & Ganss, R. Transformation of the microvascular system during multistage tumorigenesis. *Int. J. Cancer* **97**, 719-725 (2002).
3. Heldin, C. H., Rubin, K., Pietras, K. & Ostman, A. High interstitial fluid pressure — an obstacle in cancer therapy. *Nature Rev. Cancer* **4**, 806-813 (2004).
4. Jain, R. K., Tong, R. T. & Munn, L. L. Effect of vascular normalization by antiangiogenic therapy on interstitial hypertension, peritumor edema, and lymphatic metastasis: insights from a mathematical model. *Cancer Res.* **67**, 2729-2735 (2007).
5. Ganss, R., Arnold, B. & Hammerling, G. J. Mini-review: overcoming tumor-intrinsic resistance to immune effector function. *Eur. J. Immunol.* **34**, 2635-2641 (2004).



6. Buckanovich, R. J. *et al.* Endothelin B receptor mediates the endothelial barrier to T cell homing to tumors and disables immune therapy. *Nature Med.* **14**, 28–36 (2008).
7. Morikawa, S. *et al.* Abnormalities in pericytes on blood vessels and endothelial sprouts in tumors. *Am. J. Pathol.* **160**, 985–1000 (2002).
8. Song, S., Ewald, A. J., Stallcup, W., Werb, Z. & Bergers, G. PDGFR $\beta$ <sup>+</sup> perivascular progenitor cells in tumours regulate pericyte differentiation and vascular survival. *Nature Cell Biol.* **7**, 870–879 (2005).
9. Benjamin, L. E., Golijanin, D., Itin, A., Podes, D. & Keshet, E. Selective ablation of immature blood vessels in established human tumors follows vascular endothelial growth factor withdrawal. *J. Clin. Invest.* **103**, 159–165 (1999).
10. Berger, M., Bergers, G., Arnold, B., Hammerling, G. J. & Ganss, R. Regulator of G-protein signaling-5 induction in pericytes coincides with active vessel remodeling during neovascularization. *Blood* **105**, 1094–1101 (2005).
11. Ryschich, E. *et al.* Molecular fingerprinting and autocrine growth regulation of endothelial cells in a murine model of hepatocellular carcinoma. *Cancer Res.* **66**, 198–211 (2006).
12. Seaman, S. *et al.* Genes that distinguish physiological and pathological angiogenesis. *Cancer Cell* **11**, 539–554 (2007).
13. Hollinger, S. & Hepler, J. R. Cellular regulation of RGS proteins: modulators and integrators of G protein signaling. *Pharmacol. Rev.* **54**, 527–559 (2002).
14. Chen, C., Zheng, B., Han, J. & Lin, S. C. Characterization of a novel mammalian RGS protein that binds to G $\alpha$  proteins and inhibits pheromone signaling in yeast. *J. Biol. Chem.* **272**, 8679–8685 (1997).
15. Allt, G. & Lawrenson, J. G. Pericytes: cell biology and pathology. *Cells Tissues Organs* **169**, 1–11 (2001).
16. Manome, Y., Saeki, N., Yoshinaga, H., Watanabe, M. & Mizuno, S. A culture device demonstrates that hydrostatic pressure increases mRNA of RGS5 in neuroblastoma and CHC1-L in lymphocytic cells. *Cells Tissues Organs* **174**, 155–161 (2003).
17. Bondjers, C. *et al.* Transcription profiling of platelet-derived growth factor-B-deficient mouse embryos identifies RGS5 as a novel marker for pericytes and vascular smooth muscle cells. *Am. J. Pathol.* **162**, 721–729 (2003).
18. Ganss, R. & Hanahan, D. Tumor microenvironment can restrict the effectiveness of activated antitumor lymphocytes. *Cancer Res.* **58**, 4673–4681 (1998).
19. Ganss, R., Ryschich, E., Klar, E., Arnold, B. & Hammerling, G. J. Combination of T-cell therapy and trigger of inflammation induces remodeling of the vasculature and tumor eradication. *Cancer Res.* **62**, 1462–1470 (2002).
20. Garbi, N., Arnold, B., Gordon, S., Hammerling, G. J. & Ganss, R. CpG motifs as proinflammatory factors render autochthonous tumors permissive for infiltration and destruction. *J. Immunol.* **172**, 5861–5869 (2004).
21. Jain, R. K. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* **307**, 58–62 (2005).
22. Batchelor, T. T. *et al.* AZD2171, a pan-VEGF receptor tyrosine kinase inhibitor, normalizes tumor vasculature and alleviates edema in glioblastoma patients. *Cancer Cell* **11**, 83–95 (2007).
23. Winkler, F. *et al.* Kinetics of vascular normalization by VEGFR2 blockade governs brain tumor response to radiation: role of oxygenation, angiopoietin-1, and matrix metalloproteinases. *Cancer Cell* **6**, 553–563 (2004).
24. Casanovas, O., Hicklin, D. J., Bergers, G. & Hanahan, D. Drug resistance by evasion of antiangiogenic targeting of VEGF signaling in late-stage pancreatic islet tumors. *Cancer Cell* **8**, 299–309 (2005).
25. Boss, C. N. *et al.* Identification and characterization of T-cell epitopes deduced from RGS5, a novel broadly expressed tumor antigen. *Clin. Cancer Res.* **13**, 3347–3355 (2007).
26. Furuya, M. *et al.* Expression of regulator of G protein signalling protein 5 (RGS5) in the tumour vasculature of human renal cell carcinoma. *J. Pathol.* **203**, 551–558 (2004).
27. Chen, X. *et al.* Novel endothelial cell markers in hepatocellular carcinoma. *Mod. Pathol.* **17**, 1198–1210 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank G. Küblbeck, A. Klevenz, G. Hollman and S. Schmidt for technical support in establishing *Rgs5*-knockout mice, K. Bieber for assessing brain ischaemia, B. Misselwitz for providing the contrast agent Gadomer, and H. Ee and G. Bergers for critical reading of the manuscript. This study was supported by a National Health and Medical Research Council of Australia Project Grant, start-up funds from the Western Australian Institute for Medical Research and University of Western Australia (to R.G.), the Deutsche Forschungsgemeinschaft, and the EU projects MUGEN and CancerImmunoTherapy. Microscopy was carried out using facilities at the Centre for Microscopy and Microanalysis/Biomedical Image and Analysis Facility, The University of Western Australia, which are supported by University, State and Federal Government funding.

**Author Contributions** J.H. and M.M. performed animal experiments and histology, and analysed data; M.J. and F.K. performed MRI analyses; P.R. performed confocal microscopy studies; H.H.M. and T.R. coordinated and analysed brain infarct experiments; S.K. and H.-J.G. performed electron microscopy; G.J.H. and B.A. contributed to the design of *Rgs5*-knockout studies; and R.G. designed and performed experiments, coordinated all studies and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.G. ([ganss@waimr.uwa.edu.au](mailto:ganss@waimr.uwa.edu.au)).



## METHODS

**Mice and cells lines.** RIP1-Tag5 mice (on a C3H background, provided by D. Hanahan) express the oncogene Tag under the control of the rat insulin gene promoter (RIP) in pancreatic  $\beta$  cells, and develop spontaneous tumours. In RIP1-Tag5 mice, Tag is expressed in adult mice at around week 8 to 10. Knockout mice were generated on a mixed (129  $\times$  C57BL/6) background and subsequently crossed with RIP1-Tag5/C3H mice for 8 generations. Although crossings continued up to generation 12, mice from generation 8 onwards (RIP1-Tag5  $\times$  Rgs5<sup>+/−</sup>) were intercrossed with Rgs5<sup>+/−</sup> littermates to obtain larger cohorts of RIP1-Tag5 wild-type (+/+) and Rgs5-deficient (−/−) RIP1-Tag5 mice. For adoptive transfer experiments, mice transgenic for a T-cell receptor (TCR) that recognizes Tag presented by the MHC class I molecule H-2K<sup>b</sup> (referred to as TCRCD8, provided by T. Geiger and R. Flavell) or by the MHC class II molecule I-A (TagTCR1, provided by I. Förster) were used on a C3H background as previously described<sup>20</sup>. All experimental protocols were approved by the Animal Welfare Board of the Regierungspräsidium Karlsruhe, Germany, or the Animal Ethics Committee of the University of Western Australia. AG104A cells, a spontaneous fibrosarcoma of C3H mice (provided by H. Schreiber), were injected subcutaneously ( $5 \times 10^5$  cells) into C3H or Rgs5-deficient C3H mice.

**Determination of the number of angiogenic islets and tumour burden.** Angiogenic islets were isolated by retrograde perfusion after collagenase digestion. Islets with visible haemorrhaging were counted as angiogenic islets under a dissecting microscope. Tumours were microdissected from freshly excised pancreata. Tumour volumes were measured with callipers, and the formula volume =  $0.52 \times \text{width}^2 \times \text{length}$  for approximating the volume of a spheroid was applied.

**Antibodies and histological analyses.** Sections were stained with the following antibodies: anti-CD4 (rat, GK1.5,  $10 \mu\text{g ml}^{-1}$ , BD Pharmingen), anti-CD8 (rat, Ly-2,  $10 \mu\text{g ml}^{-1}$ , BD Pharmingen) and anti-Tag (rabbit polyclonal, 1:1,000, from D. Hanahan), followed by anti-rat or anti-rabbit biotinylated secondary reagents (Vector Laboratories). Anti-CD31 (rat IgG2a, MEC 13.3,  $5 \mu\text{g ml}^{-1}$ , BD Pharmingen) and anti-PDGFR $\beta$  (rat IgG2,  $10 \mu\text{g ml}^{-1}$ , eBioscience) staining was followed by cyanin 3 (Cy3) or FITC (fluorescein isothiocyanate)-conjugated IgG F(ab')<sub>2</sub> fragment goat anti-rat ( $3 \mu\text{g ml}^{-1}$ , Dianova) as the secondary reagent. Anti-NG2 (rabbit anti-mouse, 1:500, Chemicon) staining was followed by FITC-conjugated IgG immunoglobulin donkey anti-goat ( $3 \mu\text{g ml}^{-1}$ , Dianova) as secondary reagent. Stainings for anti- $\alpha$ SMA (mouse IgG1,  $5 \mu\text{g ml}^{-1}$ , Sigma) and anti-desmin (mouse IgG1, D33, 1:300, Dako) were performed using the FITC-MOM kit (Vector). The hypoxyprobe-1 kit (Chemicon) was used to detect hypoxia. Hypoxic tumour area was quantified throughout tumours on sections that were  $100 \mu\text{m}$  apart. For assessment of immune infiltration, standard histology<sup>19</sup> was performed 7 days after T-cell transfer. Tumour sections were evaluated for CD4<sup>+</sup>/CD8<sup>+</sup> lymphocytic infiltration by using a  $\times 20$  objective lens, and five independent areas were selected, digitally photographed and counted (ImagePro). Tumour tissue was counterstained with methyl green. *In situ* hybridization is described elsewhere<sup>10</sup>.

**Dynamic magnetic resonance imaging and vascular permeability.** MRI was performed using a 1.5 tesla whole-body MR-scanner (Siemens Symphony) in combination with a custom-made radio-frequency coil for excitation and signal reception. Morphologic MR-imaging was performed using a transversal T2-weighted turbo-spin echo sequence (repetition time, TR = 1,510 ms, echo time, TE = 59 ms, field of view, FOV =  $50 \times 50 \text{ mm}^2$ , matrix = 128, slice thickness = 1.0 mm). Kinetics of the contrast agent in tumours were recorded using a T1-weighted inversion-recovery Turbo FLASH (IRTF) sequence (TR = 13 ms, TE = 5.3 ms, TI = 300 ms, slice thickness = 2 mm, FOV =  $60 \times 60 \text{ mm}^2$ , Matrix 128). In total, 120 dynamic scans were acquired from two sections within

15.36 min. Five seconds after starting the dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) measurement,  $100 \mu\text{l}$  (0.1 mmol per kg body weight) of the paramagnetic contrast agent Gadomer (Bayer-Schering Pharma) were injected manually within 5 s into the tail vein. Data were analysed using the pharmacokinetic two-compartment model of Brix, providing the parameters Amplitude (related to relative blood volume) and exchange rate constant  $k_{ep}$  (surrogate marker of vessel permeability)<sup>28</sup>. Vascular permeability in AG104A mice was determined using Evans blue as described<sup>24</sup> with minor modifications. Evans blue (Sigma) was intravenously injected into tumour-bearing mice on day 15 at  $20 \text{ mg kg}^{-1}$ . Fifteen minutes later, mice were perfused with PBS, tissues excised, and Evans blue extracted from tissue for 24 h at room temperature ( $23^\circ\text{C}$ ).

**Confocal laser scanning microscopy.** Mice were injected intravenously with  $100 \mu\text{g}$  of FITC-labelled tomato lectin (from *Lycopersicon esculentum*, Vector Laboratories) in PBS. After 3 min of circulation, mice were heart-perfused with 4% paraformaldehyde in PBS. Organs were embedded in 36% albumin/3% gelatine, and  $200\text{-}\mu\text{m}$  sections were cut with a vibratome and analysed with a Biorad MRC 1000/1024 ultraviolet confocal microscope using a Nikon UV-F  $\times 20$ , NA 0.8 glycerine immersion objective or a Nikon PlanApo  $\times 60$ , NA 1.4 oil immersion objective. Images were processed by using Photoshop 8.0 (Adobe Systems).

**Transmission electron microscopy.** For electron microscopy, tumours were fixed in 4% PFA and embedded in araldite. Ultrathin sections (70 nm) were cut on an ultracut UCT Leica microtome. Sections were contrasted with uranyl-acetate and lead citrate, and analysed with an EM 900 Zeiss electron microscope.

**Brain ischaemia.** Transient cerebral ischaemia was introduced using the intraluminal filament technique<sup>29</sup>. In adult mice, the left middle cerebral artery was occluded with an 8-0 nylon filament under anaesthesia. Ninety minutes after the ischaemic insult, the filament was withdrawn. Twenty-four hours after reperfusion, the brain was removed and sections stained with silver nitrate. Normal tissue is darkly stained by silver nitrate whereas the damaged area appears bright. The infarct and oedema volumes were calculated as described<sup>30</sup> by comparing the noninfarcted right hemisphere with the left hemisphere.

**Adoptive transfers and vaccination studies.** TCRCD8 splenocytes or TagTCR1 lymph node cells were activated *in vitro* for 3 days with 10 U of recombinant human IL-2 per ml, and 25 mM Tag peptide 362–568 (SEFLIEKRI for TCRCD8 cells) or 25 nM Tag peptide 362–384 (TNRFDLLDRMDIMFGSTGSADI for TagTCR1 cells). C3H-derived splenocytes were activated with  $1 \mu\text{g ml}^{-1}$  Con A (Sigma).  $2.5 \times 10^6$  naive or activated CD4<sup>+</sup> and  $2.5 \times 10^6$  CD8<sup>+</sup> T cells were transferred. Tag was purified from High Five insect cells infected with a baculovirus expressing the SV40 early region. Mice were primed with a single subcutaneous injection (tail base) of  $50 \mu\text{g}$  Tag protein mixed with  $50 \mu\text{g}$  CpG-ODN 1668 (phosphothioate-stabilized CpG-ODN 1668, TCCATGACGTTCTGATGCT) in  $200 \mu\text{l}$  PBS. Thereafter, CpG-ODN treatment groups were intraperitoneally injected with  $50 \mu\text{g}$  Tag protein mixed with  $50 \mu\text{g}$  CpG-ODN 1668 every second week.

**Statistical analyses.** Student's *t* test (two-tailed) was used unless otherwise indicated.  $P < 0.05$  was considered to be statistically significant.

28. Brix, G. et al. Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging. *J. Comput. Assist. Tomogr.* **15**, 621–628 (1991).
29. Hata, R. et al. A reproducible model of middle cerebral artery occlusion in mice: hemodynamic, biochemical, and magnetic resonance imaging. *J. Cereb. Blood Flow Metab.* **18**, 367–375 (1998).
30. Lin, T. N., He, Y. Y., Wu, G., Khan, M. & Hsu, C. Y. Effect of brain edema on infarct volume in a focal cerebral ischemia model in rats. *Stroke* **24**, 117–121 (1993).



# 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy

Xuekui Yu<sup>1,2</sup>, Lei Jin<sup>1,2</sup> & Z. Hong Zhou<sup>1,2</sup>

Cytoplasmic polyhedrosis virus (CPV) is unique within the Reoviridae family in having a turreted single-layer capsid contained within polyhedrin inclusion bodies, yet being fully capable of cell entry and endogenous RNA transcription<sup>1–4</sup>. Biochemical data have shown that the amino-terminal 79 residues of the CPV turret protein (TP) is sufficient to bring CPV or engineered proteins into the polyhedrin matrix for micro-encapsulation<sup>5,6</sup>. Here we report the three-dimensional structure of CPV at 3.88 Å resolution using single-particle cryo-electron microscopy. Our map clearly shows the turns and deep grooves of  $\alpha$ -helices, the strand separation in  $\beta$ -sheets, and densities for loops and many bulky side chains; thus permitting atomic model-building effort from cryo-electron microscopy maps. We observed a helix-to- $\beta$ -hairpin conformational change between the two conformational states of the capsid shell protein in the region directly interacting with genomic RNA. We have also discovered a messenger RNA release hole coupled with the mRNA capping machinery unique to CPV. Furthermore, we have identified the polyhedrin-binding domain, a structure that has potential in nanobiotechnology applications.

Three-dimensional structures of several sub-viral particles in Reoviridae have been determined at atomic resolutions by X-ray crystallography<sup>7–9</sup>. Those structures have provided a wealth of information about the unity and diversity of viral organization, double-stranded (ds)RNA genome packing and transcription, and mRNA processing and releasing. For example, nascent mRNA of bluetongue viruses is capped within the capsid shell before release<sup>7,10,11</sup>, but orthoreoviruses release their nascent mRNA into a turret chamber, where it is subsequently capped<sup>8</sup>. Both of these viruses are thought to release their mRNA from the capsid shell through five-fold axial holes on icosahedral vertices. The orthoreovirus shell protein  $\lambda 1$  alone cannot assemble into icosahedral particles<sup>12</sup>. For CPV, however, previous studies have shown that the apical pentameric pore is blocked for mRNA release<sup>13</sup> and the capsid shell protein (CSP) alone can assemble into virus-like particles *in vitro*<sup>14</sup>. In addition, there is a CPV-unique haemagglutinin-like protein (that is, the 'A spike') situated on the top of each turret<sup>4</sup>. In the absence of an atomic model of the full CPV virion, many fundamental issues have remained elusive, including the structural basis for RNA packing, transcription, processing and release, and the mechanism underlying CPV's unique property of polyhedra embedding. In contrast to CPV polyhedrin protein, which forms diffraction-quality micro-crystals even *in vivo*<sup>6,15</sup>, CPV virions have eluded X-ray crystallographic efforts, probably due to the intrinsic flexibility of the haemagglutinin-like A spikes.

To address some of these issues, we attempted to push the resolution limit of single-particle cryo-electron microscopy. Focal pairs of 16-megapixel charge-coupled device (CCD) images of CPV were

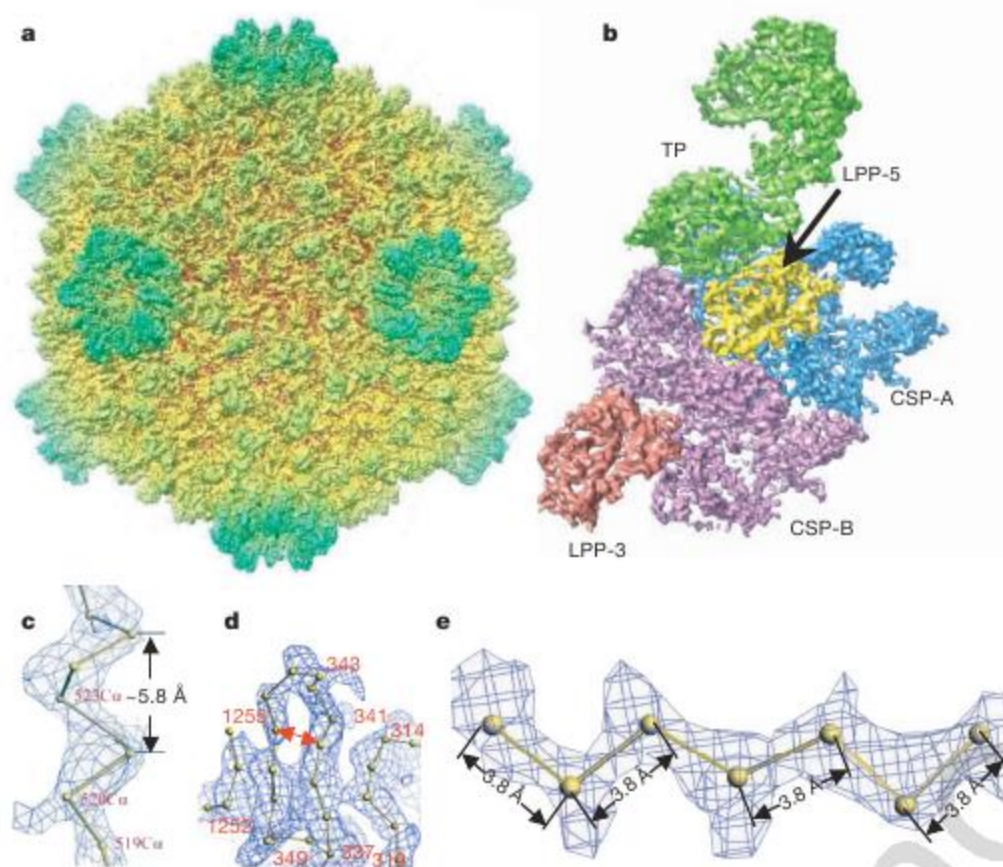
recorded at liquid-nitrogen temperature in a 300-kV FEI cryo-electron microscope (Supplementary Fig. 1a, b). Rigorous screening of images was carried out to select cryo-electron microscopy images with visible contrast transfer function (CTF) rings beyond 5 Å and without noticeable specimen drift, charging, or image astigmatism. The final map (Fig. 1a and Supplementary Movie 1) was reconstructed by merging 12,814 particles from only the close-to-focus images with a B-factor<sup>16–18</sup> of 35 Å<sup>2</sup> and underfocus values ranging from 0.15 to 1.3  $\mu$ m (Supplementary Table 1).

The effective resolution of our map was estimated to be better than 4 Å by examining the structural features revealed by the map, such as clear turns and deep grooves of helices, strand separation in  $\beta$ -sheets, a zigzagging feature of polypeptide backbones in  $\beta$ -strands, and densities of bulky side chains (Fig. 1c–e, Supplementary Fig. 2 and Supplementary Movies 2 and 3), which were comparable to those revealed in crystallographic density maps at a similar resolution<sup>19</sup>. In addition to its resolving power, the quality of the map was also convincingly verified by its accuracy, such as the pitch in  $\alpha$ -helices being  $\sim$ 5.8 Å (Fig. 1c) and the inter-consecutive C $\alpha$  distances in polypeptide chains being  $\sim$ 3.8 Å (Fig. 1e). This assessment is consistent with the 3.88 Å resolution, as estimated by the criterion of 0.5 Fourier shell correlation coefficient (Supplementary Fig. 1c). Finally, throughout the reconstruction, we only applied icosahedral averaging. Thus, the close similarities of the densities for CSP-A and CSP-B (see below), as well as the densities of the non-symmetry-related two large protrusion protein (LPP) molecules, both at an overall structural level and for a variety of detailed features, serve as an excellent internal control for the high quality of the density map (Fig. 1b and Supplementary Figs 3–5).

The 750-Å diameter CPV has a  $T = 1$  icosahedral capsid shell decorated with 12 turrets on the five-fold vertices (Fig. 1a). The asymmetric unit consists of two CSPs, two LPPs and one TP (Fig. 1b, Supplementary Fig. 3 and Supplementary Movie 4). As with other members of Reoviridae, the structures of which have been determined to atomic resolution by X-ray crystallography<sup>7–9</sup>, the CSPs of CPV have two conformational states: CSP-A and CSP-B. Although no sequence homology was detectable, our cryo-electron microscopy structures of CPV CSPs show an overall thin-plate-like shape similar to that of the orthoreovirus  $\lambda 1$  without obvious domain boundaries<sup>7,8</sup>, with the exception of the unique small protrusion domain (SPP) (Fig. 2a and Supplementary Figs 4 and 5). The identified secondary structures of CPV CSP (for example, the helices in the apical domain and the  $\beta$ -sheets in the dimerization domain) all have topological counterparts in  $\lambda 1$ , although their detail orientations differ locally, indicating that CPV CSP has a fold similar to that of  $\lambda 1$ . Using the crystallographic modelling tool O (ref. 20), we have built C $\alpha$  models for the two CSPs without the SPP (Fig. 2a and Supplementary Figs 4 and 5), two LPPs (Supplementary

<sup>1</sup>Department of Pathology and Laboratory Medicine, The University of Texas Medical School at Houston, Houston, Texas 77030, USA. <sup>2</sup>Department of Microbiology, Immunology & Molecular Genetics and The California NanoSystems Institute, University of California at Los Angeles, Los Angeles, California 90095-1594, USA.





**Figure 1 | Overall structure of the CPV capsid.**

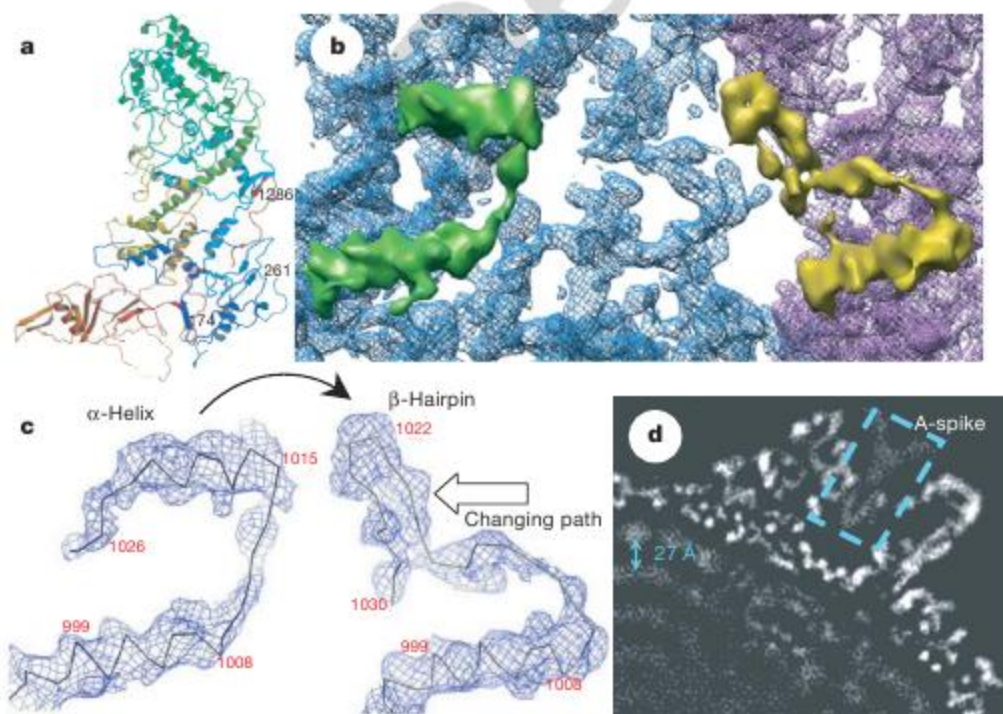
**a**, Radially coloured, shaded surface view of the CPV reconstruction as viewed along a two-fold axis. **b**, An extracted asymmetric unit, colour coded by protein subunits, including one turret protein (TP) (green), two copies of the capsid shell proteins (CSP-A in blue and CSP-B in purple), and two copies of the large protrusion protein (LPP-5, near the five-fold axis, in yellow; LPP-3, near the three-fold axis, in brown).

**c, d**, Views of density maps of one  $\alpha$ -helix (**c**) and four  $\beta$ -strands (**d**) for CSP-B superimposed with the corresponding  $C\alpha$  model, showing the clear turn and deep groove of the  $\alpha$ -helix with a pitch of  $\sim 5.8$  Å and the density for the bulky side chain of residue 520 in **c**, and the clear separation of  $\beta$ -strands in **d** (note also the densities for the bulky side chains of amino acids 341 and 343). **e**,  $C\alpha$  model of a  $\beta$ -strand of TP in the density map, showing zigzagging of the backbone and the clear densities for side chains. All the inter-consecutive  $C\alpha$  distances are  $\sim 3.8$  Å, which is a well established distance in polypeptide chains.

Fig. 6), and two N-terminal domains of TP based on our density map. At the current resolution, accurate registration of residues remains challenging and, consequently, residue numberings were tentative.

A marked local conformational change has been identified between the two conformers CSP-A and CSP-B in their regions interacting with viral RNA (Fig. 2), in addition to a conserved global conformational swing about a pivot in the middle<sup>7,8</sup> (Supplementary Figs 4 and 5). Whereas residues 1014–1024 form an  $\alpha$ -helix in CSP-A, the same fragment exists as part of a  $\beta$ -hairpin (residues 1017–1029) in CSP-B. The  $\alpha$ -helix of CSP-A is connected to another approximately parallel  $\alpha$ -helix (residues 996–1009) through a 4-residue loop (residues 1010–1013). In CSP-B, this loop is rotated about

45° clockwise (viewed from inside the capsid) and subsequently pushes the connected  $\beta$ -hairpin to rotate  $\sim 90^\circ$  clockwise and backward  $\sim 20$  Å to the edge of the carapace domain to interact with its adjacent CSP-A (Fig. 2b, c). The genomic dsRNA is coiled and tightly packed within the capsid with a multilayer arrangement<sup>21</sup>. The outermost RNA layer interacts with the inner side of the capsid shell (Fig. 2d and Supplementary Fig. 7). The  $\alpha$ -helix (residues 1014–1024) of CSP-A is the most inward component and most likely to be involved in the interaction with viral RNA. The functional implications of the observed conformational change remain unknown; however, the interactions with viral RNA genome suggest an association with RNA packaging, replication and transcription (See 'Discussion' in the Supplementary Information).



**Figure 2 | A conformational change between CSP-A and CSP-B: implication for packing and sliding of the dsRNA genome.**

**a**, Atomic model of CSP-A, coloured from blue at the N terminus to red at the C terminus. **b**, Close-up view of the CSP-A (blue) and CSP-B (purple) density map, showing that an  $\alpha$ -helix (the upper helix within the green density) in CSP-A transforms into part of a  $\beta$ -hairpin in CSP-B (the upper part in the yellow density). **c**, Density maps with  $C\alpha$  models, showing the conformational change between CSP-A (left) and CSP-B (right). One  $\alpha$ -helix in CSP-A transforms into part of the  $\beta$ -hairpin in CSP-B (indicated by curved arrow). The changing path is also indicated by an empty arrow. **d**, A 10 Å slab extracted from the two-fold map showing the ordered dsRNA genome with a  $\sim 27$  Å distance between the adjacent dsRNA strands (arrow). The dotted box indicates the A-spike plug all the way in the central chamber of the turret.



Uncapped nascent mRNA is unstable and degrades prematurely if released into the cytoplasm. Viruses in the Reoviridae overcome this problem by capping nascent mRNAs before releasing them. Whereas viruses with non-turreted cores cap their mRNA within the capsid shell, turreted viruses cap their mRNA in the turret outside the capsid shell. CPV TP differs from orthoreovirus  $\lambda 2$  despite sharing similar domain architectures. The  $\lambda 2$  pentamer of orthoreovirus is a hollow cylinder with five flaps forming a lid on the top<sup>8</sup>. The TP pentamer of CPV is a hollow cone with a wide opening on the top, as TP lacks the immunoglobulin fold flap domains of the orthoreovirus  $\lambda 2$ . Instead, the central chamber of the CPV turret is plugged all the way down by its haemagglutinin-like protein A-spike (dotted box in Fig. 2d and yellow in Supplementary Fig. 8a, b and Supplementary Movie 6), which confers the CPV's infectivity. The A-spike adopts a shape complementary to that of the inner space of the turret and totally blocks the central channel at the lower part of the turret, leaving only a peripheral space of  $\sim 20$  Å in the middle to upper regions of the turret (Fig. 2d, Supplementary Fig. 8a and Supplementary Movie 6). In keeping with this arrangement of the turret, the CSPs of CPV do not form a single hole at each five-fold axis for mRNA release, like orthoreovirus<sup>8</sup>, but instead form five holes surrounding the five-fold axis, similar to those seen in a low-resolution structure of the rotavirus<sup>22</sup>. In CPV, each of those holes is  $\sim 10$  Å in diameter, formed by the apical domains of two CSP-A's and one CSP-B (Fig. 3a), and is immediately adjacent to the active-site cleft of the guanylyltransferase (GTase) of TP (Fig. 3b, c and Supplementary Fig. 8c). Notably, the narrowest region of each hole is formed by three loops (Fig. 3a).

Consistent with this difference in the location of the putative mRNA release hole, the structure of the CPV GTase domain, which

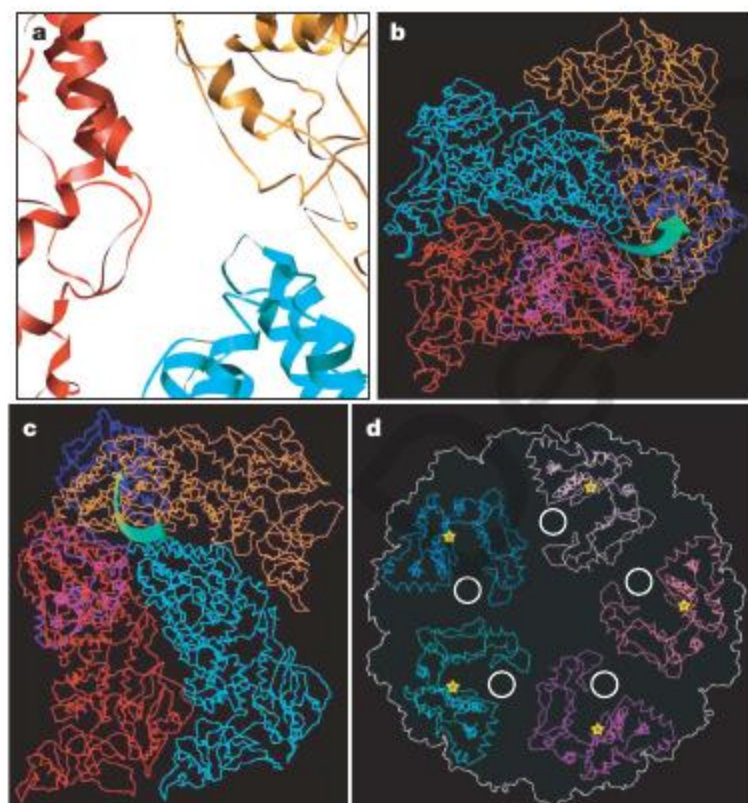
catalyses the first step of RNA capping, also differs from that of orthoreovirus. Although CPV and orthoreovirus belong to the same family, the CPV GTase domain is rather like the GTase of the completely unrelated chlorella virus PBCV-1 (ref. 23) in having two sub-domains, forming a deep and narrow cleft (Fig. 3d and Supplementary Fig. 9). On the basis of the structural resemblance, we speculate that the active site of the CPV GTase is on the larger sub-domain, opening to the cleft. Rather than facing the central chamber, as for the orthoreovirus GTase, the active-site cleft of the CPV GTase rotates away from the central chamber, thus facing sideways towards the putative RNA release hole. This configuration of GTase allows it to be coupled with the mRNA release hole in CPV (Fig. 3b–d). This configuration is different from that of orthoreovirus as mentioned above, but still ensures that the nascent mRNA bumps into the active-site cleft of GTase immediately after it exits the capsid shell, resulting in immediate and efficient capping as in the orthoreovirus (Fig. 3b–d and Supplementary Movie 7). The PBCV-1 GTase exists in two conformations: the inactive open conformation and the active closed conformation<sup>23</sup>. Notably, the CPV GTase domain resembles the inactive open conformation more closely, consistent with the fact that our structure was obtained from quiescent, non-transcribing CPV virions.

Mutagenesis studies have shown that the N-terminal 79 residues of TP are required for specific interaction of the CPV capsid with polyhedrin trimer during polyhedra-embedding or micro-encapsulation into polyhedra<sup>5,6,24</sup>. Consistent with this observation, our density map of TP shows a CPV-unique domain anchored to the N-terminal end of the GTase domain through a clearly resolved linker (Fig. 4 and Supplementary Movies 8 and 9). This domain has a palm-shaped structure with a dimension of 15 Å by 20 Å consisting of four  $\beta$ -strands (Fig. 4d, e). Because the GTase domain in orthoreovirus is the N-terminal end of the capping complex  $\lambda 2$ , the additional domain attached to the CPV GTase domain is probably the N-terminal domain of CPV TP (that is, the CPV-unique polyhedrin-binding domain (PBD)). Indeed, our model shows that this domain is  $\beta$ -strand rich, which is in agreement with the secondary structure prediction from the primary structure of the TP N-terminal 79 amino acids. This domain is located on the outer surface of the CPV turret, and is thus easily accessible by the polyhedrin trimer. Its sheet-like feature apparently provides a large interaction surface, which may be the structural basis for its strong and specific binding to the polyhedrin trimer<sup>6</sup>.

Because the PBD has the unique ability to bring proteins of various sizes into nano-scale inclusion bodies (polyhedra) with controllable disassembly behaviour, this would offer better systems for delivery and protein microarrays<sup>5</sup>. Our elucidation of the structural details of this domain opens doors for rational designs of such systems.

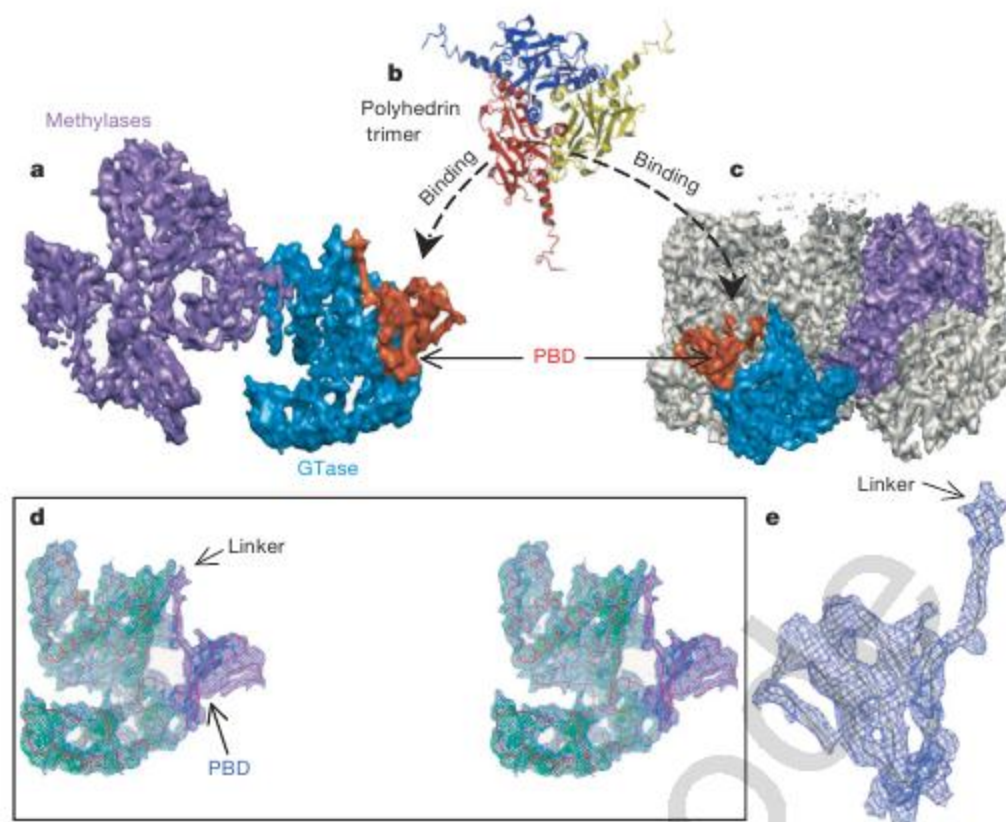
We have obtained a 3.88 Å density map of CPV by single-particle cryo-electron microscopy, which is of sufficient quality for us to attempt atomic modelling and has led to three important findings (summarized in Supplementary Fig. 10). In the last decade there have been great efforts at sub-nanometre resolution in single-particle cryo-electron microscopy reconstructions<sup>16,25</sup>. Our structure was obtained from only  $\sim 13,000$  particles recorded in a 300-kV microscope without an energy filter. With further advancement of cryo-electron microscopy instrumentation, the development of more sophisticated algorithms and ever-growing computational power, there is every reason to believe that we are entering the age of atomic resolution for single-particle cryo-electron microscopy, which will have profound impacts on biology, chemistry and medicine.

**Note added in proof:** Recently, three other papers have appeared describing near-atomic-resolution structures of the rotavirus double-shell particles<sup>26</sup>, the  $\phi 15$  bacteriophage<sup>27</sup> and the GroEL complex<sup>28</sup>. These advancements illustrate the potential towards atomic-resolution structural determination by single-particle cryo-electron microscopy.



**Figure 3 | Nascent mRNA release hole coupled with the GTase active site of TP in a way unique to CPV.** **a**, Close-up view of the atomic model of CSP-A (two copies) and CSP-B. The mRNA release hole comprises loops contributed by the two copies of CSP-A (red and orange) and one CSP-B (cyan). **b**, **c**, Cx models of one CSP-B (cyan), two copies of CSP-A (red and orange) and two GTase domains (blue and light blue), showing the mRNA releasing and capping pathway, which is illustrated by green arrows in **b** (viewed from outside of the capsid) and **c** (viewed from inside of the capsid). **d**, Cx models of the GTase domains in a turret, showing that the active-site cleft (yellow stars) of GTase rotates away from the central chamber, facing sideways, thus coupled with the mRNA release hole (white circle).





**Figure 4 | Location and structure of the unique polyhedrin-binding domain (PBD) of TP.** **a**, Shaded surface representation of TP viewed from inside, showing the two methylase domains (purple), GTase domain (blue), and CPV's unique PBD (orange). The PBD is absent from other dsRNA viruses and consists of residues 1–79 of TP. **b**, Ribbon model of the CPV polyhedrin trimer, which is the building block of the crystalline polyhedra<sup>6</sup>. **c**, View of

one entire turret, showing the orientation and position of one TP (coloured) in the turret. **d**, Stereo view of a cryo-electron microscopy density map superimposed with the C $\alpha$  model of GTase domain (green) and PBD (blue), which are connected together through a linker. **e**, Blown-up view of the model of PBD superimposed with its cryo-electron microscopy density.

## METHODS SUMMARY

The CPV virions were isolated and purified from infected *Bombyx mori* larvae using a previously described procedure<sup>4,13</sup>, with some modifications primarily to preserve the structural integrity for high-resolution structure determination. Focal-pair images were collected at liquid-nitrogen temperature on a 16-megapixel CCD camera (TVIPS) in an FEI Polara cryo-electron microscope operated at 300 kV with a dosage of  $\sim 20\text{ e}^-$  per  $\text{\AA}^2$  at a magnification of  $\times 154,380$  (step size:  $0.97\text{ \AA}$  per pixel). Determination of particle orientation, centre parameters and subsequent three-dimensional reconstruction were carried out using the IMIRS package<sup>29</sup>, using integrated strategies of common lines, projection matching<sup>30</sup> and Fourier-Bessel synthesis methods<sup>31,32</sup>. Before the merging of particles for three-dimensional reconstruction, the Fourier transform values of individual images were corrected for the CTF with 15% amplitude contrast and a decay factor of  $35\text{ \AA}^2$ . We estimated the effective resolution of the final reconstruction to be beyond  $4\text{ \AA}$  using a structure-feature-based criterion commonly accepted in protein crystallography, such as visible turns and deep grooves for helices, strand separations in  $\beta$ -sheets and clear densities for bulky side chains. This is consistent with the  $3.88\text{ \AA}$  effective resolution based on the criterion of the 0.5 Fourier shell correlation coefficient between two reconstructions. Because CPV has no recognizable sequence homologies (that is,  $<8\%$  sequence identity) with orthoreovirus<sup>6</sup>, bluetongue virus<sup>7</sup> and phytoreovirus<sup>9</sup>, experimental C $\alpha$ -trace model building based on cryo-electron microscopy densities was necessary and was accomplished using O (ref. 20).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 October 2007; accepted 3 March 2008.

Published online 30 April 2008.

- Mertens, P. P. C., Attoui, H., Duncan, R. & Dermody, T. S. in *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses* (eds Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A.) 447–454 (Elsevier/Academic Press, London, 2005).
- Zhou, Z. H. in *Segmented Double-Stranded RNA Viruses: Structure and Molecular Biology* (ed. Patton, J. T.) 27–43 (Caister Academic Press, Norfolk, 2008).
- Hill, C. L. *et al.* The structure of a cypovirus and the functional organization of dsRNA viruses. *Nature Struct. Biol.* **6**, 565–568 (1999).

- Zhang, H. *et al.* Visualization of protein-RNA interactions in cytoplasmic polyhedrosis virus. *J. Virol.* **73**, 1624–1629 (1999).
- Ikeda, K. *et al.* Immobilization of diverse foreign proteins in viral polyhedra and potential application for protein microarrays. *Proteomics* **6**, 54–66 (2006).
- Coulbaly, F. *et al.* The molecular organization of cypovirus polyhedra. *Nature* **446**, 97–101 (2007).
- Grimes, J. M. *et al.* The atomic structure of the bluetongue virus core. *Nature* **395**, 470–478 (1998).
- Reinisch, K. M., Nibert, M. L. & Harrison, S. C. Structure of the reovirus core at  $3.6\text{ \AA}$  resolution. *Nature* **404**, 960–967 (2000).
- Nakagawa, A. *et al.* The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* **11**, 1227–1238 (2003).
- Gouet, P. *et al.* The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. *Cell* **97**, 481–490 (1999).
- Diprose, J. M. *et al.* Translocation portals for the substrates and products of a viral transcription complex: the bluetongue virus core. *EMBO J.* **20**, 7229–7239 (2001).
- Xu, P., Miller, S. E. & Joklik, W. K. Generation of reovirus core-like particles in cells infected with hybrid vaccinia viruses that express genome segments L1, L2, L3, and S2. *Virology* **197**, 726–731 (1993).
- Zhou, Z. H., Zhang, H., Jakana, J., Lu, X.-Y. & Zhang, J.-Q. Cytoplasmic polyhedrosis virus structure at  $8\text{ \AA}$  by electron cryomicroscopy: structural basis of capsid stability and mRNA processing regulation. *Structure* **11**, 651–663 (2003).
- Hagiwara, K. & Naitow, H. Assembly into single-shelled virus-like particles by major capsid protein VP1 encoded by genome segment S1 of *Bombyx mori* cypovirus 1. *J. Gen. Virol.* **84**, 2439–2441 (2003).
- Di, X., Sun, Y.-k., McCrae, M. A. & Rossmann, M. G. X-ray powder pattern analysis of cytoplasmic polyhedrosis virus inclusion bodies. *Virology* **180**, 153–158 (1991).
- Böttcher, B., Wynne, S. A. & Crowther, R. A. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88–91 (1997).
- Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
- Zhou, Z. H. & Chiu, W. Determination of icosahedral virus structures by electron cryomicroscopy at subnanometer resolution. *Adv. Protein Chem.* **64**, 93–124 (2003).
- Inaba, K. *et al.* Crystal structure of the DsbB-DsbA complex reveals a mechanism of disulfide bond generation. *Cell* **127**, 789–801 (2006).



20. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
21. Xia, Q., Jakana, J., Zhang, J. Q. & Zhou, Z. H. Structural comparisons of empty and full cytoplasmic polyhedrosis virus: protein-RNA interactions and implications for endogenous RNA transcription mechanism. *J. Biol. Chem.* **278**, 1094–1100 (2003).
22. Lawton, J. A., Estes, M. K. & Prasad, B. V. Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles. *Nature Struct. Biol.* **4**, 118–121 (1997).
23. Hakansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. X-ray crystallography reveals a large conformational change during guanylyl transfer by mRNA capping enzymes. *Cell* **89**, 545–553 (1997).
24. Ikeda, K. et al. Molecular characterization of *Bombyx mori* cytoplasmic polyhedrosis virus genome segment 4. *J. Virol.* **75**, 988–995 (2001).
25. Conway, J. F. et al. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* **386**, 91–94 (1997).
26. Zhang, X. et al. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc. Natl Acad. Sci. USA* **105**, 1867–1872 (2008).
27. Jiang, W. et al. Backbone structure of the infectious  $\alpha$ 15 virus capsid revealed by electron cryomicroscopy. *Nature* **451**, 1130–1134 (2008).
28. Ludtke, S. J. et al. *De novo* backbone trace of GroEL from single particle electron cryomicroscopy. *Structure* **16**, 441–448 (2008).
29. Liang, Y., Ke, E. Y. & Zhou, Z. H. IMIRS: a high-resolution 3D reconstruction package integrated with a relational image database. *J. Struct. Biol.* **137**, 292–304 (2002).
30. Baker, T. S. & Cheng, R. H. A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J. Struct. Biol.* **116**, 120–130 (1996).
31. Crowther, R. A., Amos, L. A., Finch, J. T., DeRosier, D. J. & Klug, A. Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* **226**, 421–425 (1970).
32. Crowther, R. A. Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Phil. Trans. R. Soc. Lond. B* **261**, 221–230 (1971).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research is supported in part by grants from NIH and the Welch Foundation. We are grateful to W. Chiu for his advice and support of this project. We thank Y. Liang, J. Jakana, M. Baker and W. Chiu for their participation at the preliminary stage of this project; J.-Q. Zhang for providing the CPV-containing polyhedra sample; I. Atanasov for assistance during cryo-electron microscopy imaging; X. Zhang for graphics illustration; and P. Lo for reading our manuscript.

**Author Contributions** X.Y. and Z.H.Z. collected the cryo-electron microscopy data; X.Y. processed the data; L.J. built the models; all authors participated in the structure interpretation and manuscript preparation.

**Author Information** The cryo-electron microscopy density map has been deposited to the EM Data Bank with accession codes EMD-1508. Coordinates for CSP-A, CSP-B and the N-terminal two domains of TP have been deposited in the Protein Data Bank with accession code 3CNF. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Z.H.Z. (Hong.Zhou@UCLA.edu).



## METHODS

**Virus purification.** In contrast to the stringent requirement on sample purity by conventional X-ray crystallography and nuclear magnetic resonance (NMR), we believe that sample integrity is more important than purity for high-resolution single-particle cryo-electron microscopy structure determination of viruses and large protein assemblies. In cryo-electron microscopy images, impurities or contaminations, if ever existing, can be distinguished from particles of interest either visually or by computational means, thus they will not have a severe impact on subsequent data processing. However, subtle structural deformation or flexibility of the complexes cannot be easily sorted out and can become a limiting factor towards high-resolution reconstruction. To avoid possible structural damages caused by multi-step sucrose-gradient centrifugation, we isolated the CPV virions from *Bombyx mori* larvae using a simplified protocol modified from a previous procedure<sup>4,13</sup>. Purified polyhedra were treated with an alkaline solution of 0.2 M Na<sub>2</sub>CO<sub>3</sub>-NaHCO<sub>3</sub> (pH 10.8) for 1 h. The suspension was centrifuged at 10,000g for 40 min. The supernatant was collected and centrifuged at 80,000g for 60 min at 4 °C to pellet the CPV virions. The pellet was then directly re-suspended in 10 mM PBS (pH 7.4) and immediately flash-frozen for cryo-electron microscopy imaging. We eliminated the subsequent sucrose gradient centrifugation steps used in previous protocols<sup>4,13</sup>.

**Cryo-electron microscopy imaging.** The CPV virions were embedded in a thin layer of vitreous ice suspended across the holes of holey carbon films for cryo-electron microscopy imaging. The images were collected at liquid-nitrogen temperature in a 300-kV FEI Polara cryo-electron microscope with a dosage of ~20 e<sup>-</sup> per Å<sup>2</sup> at a magnification of 154,380 and a step size of 0.97 Å per pixel. Focal pairs of micrographs were recorded on a 16-megapixel CCD camera (TVIPS) using the EMMENU software. The first close-to-focus micrographs in the focal pairs were aimed at 0.5–1.0 µm underfocus. The second micrographs were taken at 1.5 µm farther underfocus and were used only for the determination of preliminary orientation parameters. The defocus values were estimated from the positions of the CTF rings in the incoherently averaged Fourier transforms of particle images. As compared to conventional imaging on photographic films, imaging on CCD exposes a much smaller sample area, thus reducing possible beam-induced specimen movement/charging and the effect of defocus variation within each image. The latter allows a more accurate defocus value determination.

**Three-dimensional reconstruction and visualization.** Determination of particle orientation, centre parameters and subsequent three-dimensional reconstruction were carried out using the IMIRS package<sup>29</sup>, which uses integrated approaches of common lines<sup>32</sup>, projection matching<sup>30</sup>, Fourier-Bessel synthesis<sup>31,32</sup> and spherical harmonics reconstruction<sup>33</sup> methods. Before merging of particles for three-dimensional reconstruction, the Fourier transform values of individual images were corrected for the CTF with 15% amplitude contrast and a decay factor of 35 Å<sup>2</sup>. Reconstructions were performed independently from images recorded in individual cryo-electron microscopy imaging sessions and were evaluated based on resolvability of high-resolution structural features. Many data sets were discarded owing to difficulties in pushing beyond 5 Å resolution, and the final reconstruction was only from images selected from two best cryo-electron microscopy sessions. The images from these two extended sessions were independently processed to 5 Å and combined for further

refinement. The final refined data set was split into two halves for Fourier shell correlation calculation.

The three-dimensional reconstruction was visualized using Chimera<sup>34</sup>. The density maps were displayed by contouring at 1.5σ (standard deviation) above the average density values unless otherwise specified. Each molecular component was manually segmented out using Chimera. Molecular boundaries were well delineated by interactive three-dimensional examination of the continuity of mass densities by varying the contouring levels (1.5σ–3.0σ).

**Atomic model building and visualization.** The sequences of CPV proteins are highly diversified, as no significant sequence homologies have been detected (<8% sequence identity). So far, we have not been able to perform homology modelling for CSP, LPP and TP, despite extensive efforts using a variety of bioinformatics tools. Interpretation of the final experimental density map and model building into the densities were performed using the software O (ref. 20). To facilitate polypeptide chain tracing, the density map was skeletonized in O. For CSP-A and CSP-B, Cα models were built in O according to the skeleton and following the overall topology of orthoreovirus λ1 proteins<sup>8</sup>. For the GTase domain of the TP, tentative Cα tracing was based on the density map only. Subsequently, a poly-Ala model was built from the Cα model using the `lego_auto_MC` command in O. Fitting of the model to the map and adjusting of clashing atoms were done manually in O. Local stereochemistry was improved using the `real-space refinement` (the `refine_zone` command) in O. Tentative registration of amino acids was based on the density map, the secondary structure profiles predicted from the primary amino-acid sequences and the overall topology. The current model contains residues 1–26, 48–275 and 291–391 of the 1,057 amino acid residues for TP; residues 174–255, 261–720, 889–1095, 1098–1214, 1220–1255 and 1265–1287 of the 1,333 residues for CSP-A; and residues 174–255, 261–720, 889–1214, 1222–1255 and 1269–1285 of the 1,333 residues for CSP-B. This accounts for a total of 2,195 amino acids for the icosahedral asymmetric unit.

Secondary structures were clearly recognizable for the small protrusion domain of CSP, and the two methylase domains of TP, even though their backbones could not be traced from the map, probably due to their inherent flexibility. For LPP, its identity and sequence are still ambiguous. One possible candidate for LPP is the VP5 protein; however, the predicted molecular weight based on the sequence of VP5 is much larger than the density for LPP, probably as a result of proteolysis<sup>35</sup>. Therefore, those domains or molecules were not included in the current model.

Figures and Supplementary Movies were prepared using Chimera<sup>34</sup>, O (ref. 20), Molscript<sup>36</sup> and Pymol<sup>37</sup>.

33. Liu, H. et al. Symmetry-adapted spherical harmonics method for high-resolution 3D single-particle reconstructions. *J. Struct. Biol.* **161**, 64–73 (2008).
34. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
35. Hagiwara, K. & Matsumoto, T. Nucleotide sequences of genome segments 6 and 7 of *Bombyx mori* cypovirus 1, encoding the viral structural proteins V4 and V5, respectively. *J. Gen. Virol.* **81**, 1143–1147 (2000).
36. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
37. DeLano, W. L. *The PyMOL User's Manual* (DeLano Scientific, Palo Alto, California, 2002).



# Crystal structure of squid rhodopsin

Midori Murakami<sup>1</sup> & Tsutomu Kouyama<sup>1,2</sup>

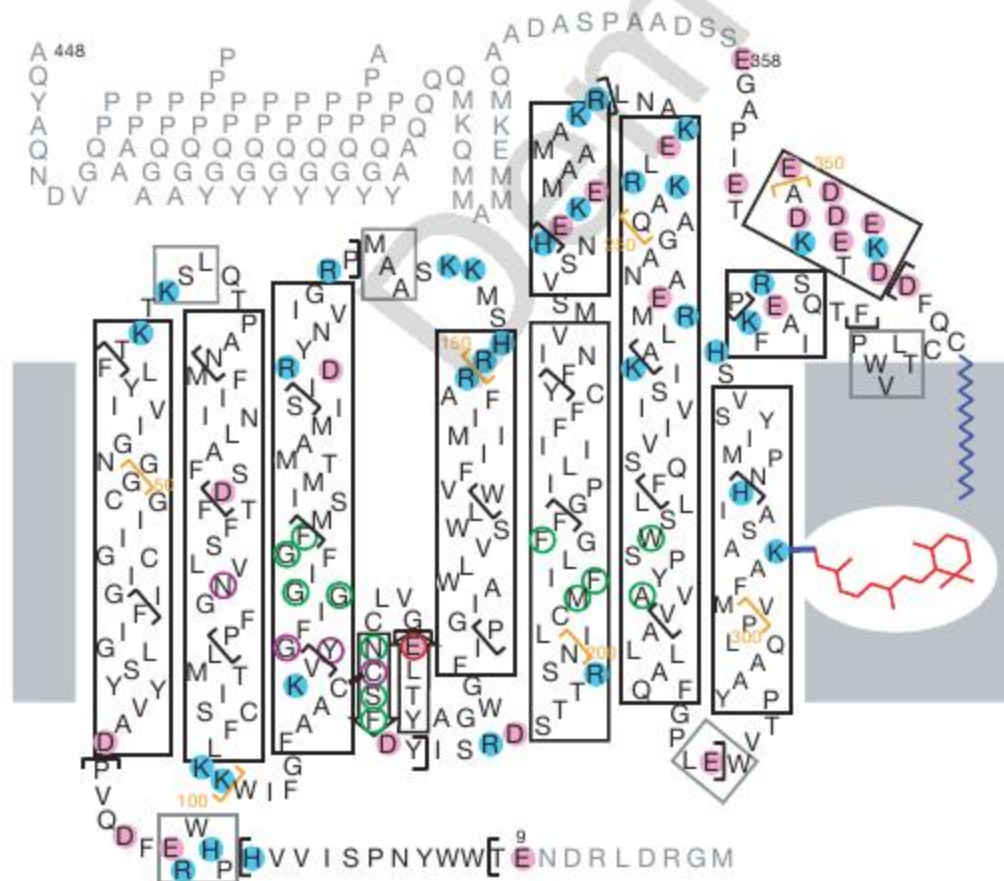
Invertebrate phototransduction uses an inositol-1,4,5-trisphosphate signalling cascade in which photoactivated rhodopsin stimulates a G<sub>q</sub>-type G protein, that is, a class of G protein that stimulates membrane-bound phospholipase C $\beta$ . The same cascade is used by many G-protein-coupled receptors, indicating that invertebrate rhodopsin is a prototypical member. Here we report the crystal structure of squid (*Todarodes pacificus*) rhodopsin at 2.5 Å resolution. Among seven transmembrane  $\alpha$ -helices, helices V and VI extend into the cytoplasmic medium and, together with two cytoplasmic helices, they form a rigid protrusion from the membrane surface. This peculiar structure, which is not seen in bovine rhodopsin, seems to be crucial for the recognition of G<sub>q</sub>-type G proteins. The retinal Schiff base forms a hydrogen bond to Asn 87 or Tyr 111; it is far from the putative counterion Glu 180. In the crystal, a tight association is formed between the amino-terminal polypeptides of neighbouring monomers; this intermembrane dimerization may be responsible for the organization of hexagonally packed microvillar membranes in the photoreceptor rhabdom.

G-protein-coupled receptors (GPCRs) are activated by a variety of external physical and chemical stimuli and they transmit the stimuli to intracellular signalling cascades by activation of G proteins at the cytoplasmic surface<sup>1</sup>. Rhodopsin is the primary photoreceptor molecule in the visual signalling cascade<sup>2–4</sup>. Like other visual pigments, invertebrate rhodopsin contains 11-*cis* retinal attached to a lysine residue, the photoisomerization of which results in the activation of a heterotrimeric G protein (guanine-nucleotide-binding protein). In contrast to vertebrate vision, in which signal transduction is mediated by the second messenger cyclic GMP, invertebrate phototransduction uses an inositol-1,4,5-trisphosphate signalling cascade in which a

G<sub>q</sub>-type G protein is stimulated by photoactivated rhodopsin<sup>5</sup>. As the latter is used by numerous GPCRs, invertebrate rhodopsin can be regarded as a prototypical member of the GPCR family.

## Crystal structure of squid rhodopsin

Squid (*Todarodes pacificus*) rhodopsin contains 448 amino acids with a molecular mass of ~50 kDa<sup>6</sup>. Its polypeptide chain is ~100 residues longer than those of vertebrate counterparts, primarily due to a 10-kDa carboxy-terminal extension containing proline-rich sequences (Fig. 1). This C-terminal extension is important in the regulation of intracellular trafficking and photoreceptor morphogenesis<sup>7</sup>, but its



**Figure 1 | Schematic diagram of the topology of squid rhodopsin.**  $\alpha$ -Helices are denoted by black rectangles, 3<sub>10</sub> helices as grey rectangles and  $\beta$ -strands as open arrows. In the current structural model, the polypeptide chain is traced from Glu 9 to Glu 358 (shown by black letters). Grey letters indicate residues that are truncated or disordered. Retinal bound to Lys 305 and palmitoyl bound to Cys 337 are included in the structural model. A disulphide bond is formed between Cys 108 and Cys 186. The counterion Glu 180, residues within 3.5 Å of the Schiff base and within 4 Å of the retinal polyene chain are circled by red, purple and green lines, respectively. Positively and negatively ionizable residues are marked with blue and pink circles, respectively.

<sup>1</sup>Department of Physics, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan. <sup>2</sup>RIKEN Harima Institute/SPRING-8, 1-1-1, Kouto, Sayo, Hyogo 679-5148, Japan.



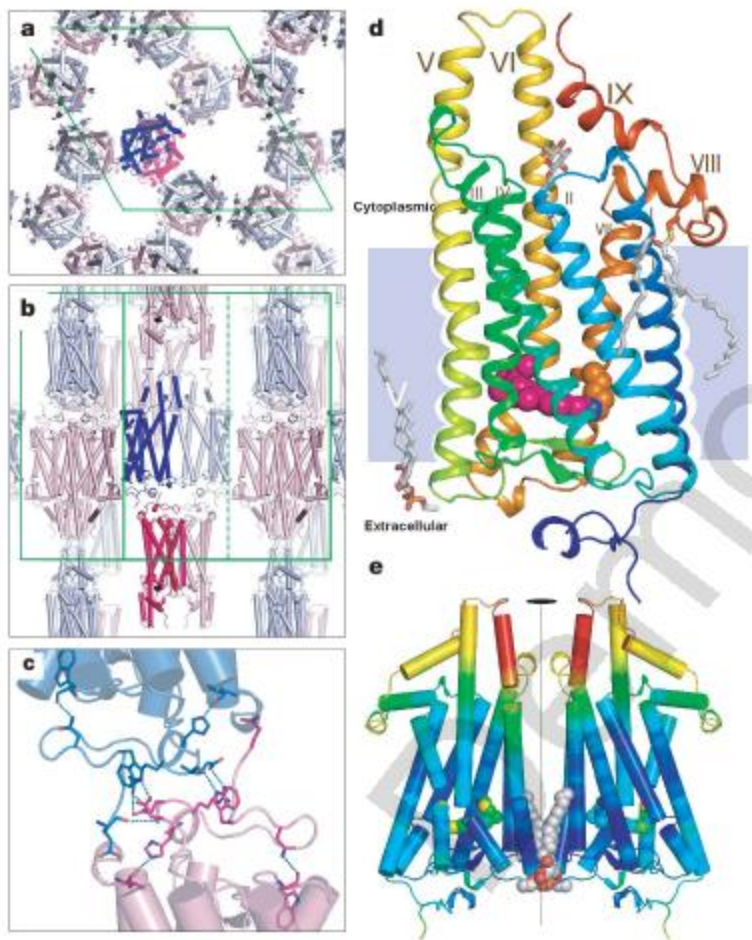
deletion does not affect the ability of rhodopsin to activate G proteins<sup>8</sup>. In this study, the C-terminal extension was removed by cleaving a peptide bond at Glu 373 (or Glu 358) with V8-protease, and C-terminally truncated rhodopsin was crystallized according to a procedure described previously<sup>9,10</sup>. Briefly, the truncated rhodopsin was extracted selectively from the microvillar membranes with octylglucoside in the presence of zinc acetate and it was crystallized into the hexagonal  $P6_2$  crystal.

In this crystal form, the asymmetric unit contains two molecules with similar structures. In Fig. 2a, two adjacent subunits that are related by a local two-fold axis lying in the  $ab$  plane are drawn in red and blue. The N-terminal peptides of these subunits are associated tightly with each other by salt bridges between Arg 24 and Glu 25 of both subunits (Fig. 2c). It is feasible that the same type of protein–protein contact exists between adjacent microvillar membranes, contributing to the membrane organization into the hexagonally packed cylinders that are observed in the photoreceptor rhabdom<sup>11</sup>. In the crystal, another type of protein association is seen between two adjacent B-subunits that are related by a crystallographic two-fold axis (Fig. 2e). In this dimeric association, the

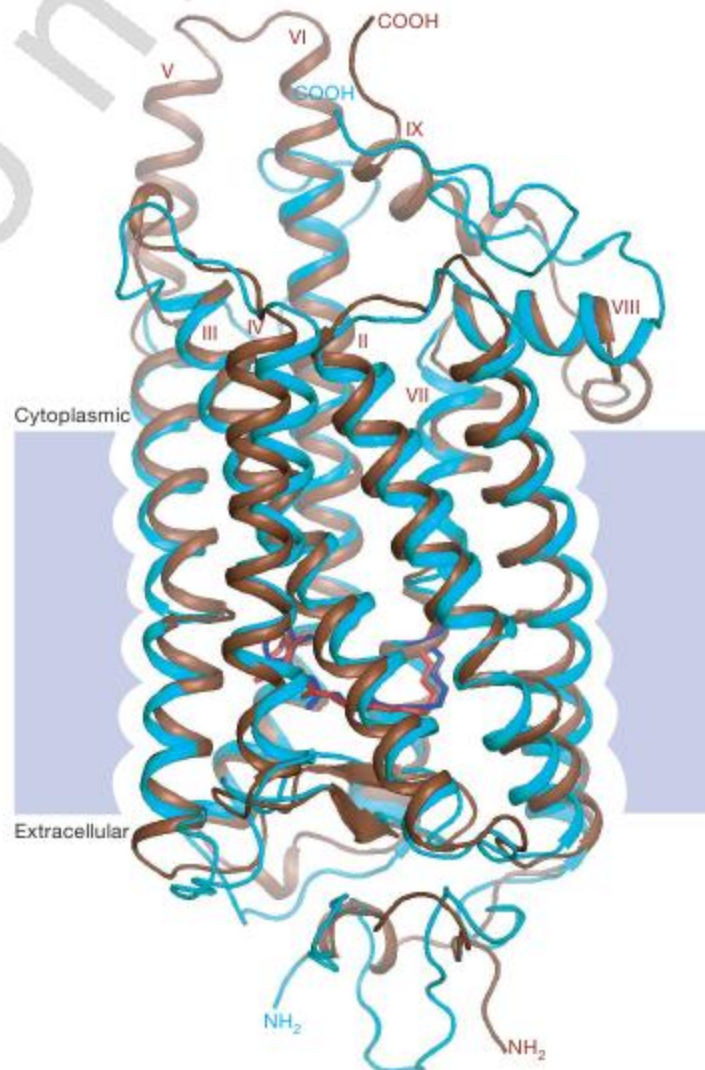
protein–protein interactions are reinforced by phospholipids bound to intradimer crevices. This dimeric structure is similar in geometry to a dimer model that was recently proposed to explain atomic force micrographs of rat disc membranes<sup>12</sup>. However, as two adjacent A-subunits are differently arranged around the crystallographic two-fold axis, it is possible that the protein association within the same membrane is weaker than the intermembrane protein–protein association.

### Overall structure

C-terminally truncated rhodopsin, the polypeptide of which is traced from Glu 9 to Glu 358, is composed of seven transmembrane helices (from I to VII) and two cytoplasmic helices (VIII and IX) (Fig. 2d). Notably, helices V and VI protrude into the cytoplasmic medium, forming a column projecting 25 Å from the membrane surface. The presence of this column has been implicated by a low-resolution electron-density map of squid rhodopsin in a two-dimensional crystal<sup>13</sup>. Helix V is divided into a membrane-embedded region and a medium-exposed region; owing to a flexible joint at Ser 226, the latter helical part has a large motional freedom (Fig. 2e). In contrast, helix VI extends into the medium as a rigid spiral and its hydrophilic extension interacts with the C-terminal end of helix IX. Although helix IX contains many acidic residues, its motional freedom is suppressed by interactions with helix VIII, which is anchored to the cytoplasmic membrane surface by a palmitoyl group attached to Cys 337. Such an organized protein folding in the cytoplasmic



**Figure 2 | Crystal structure of C-terminally truncated squid rhodopsin.** **a, b**, Crystal packing viewed along the  $c$ -axis and  $a^*$ -axis, respectively. The subunits drawn in blue (A-subunit) and magenta (B-subunit) form a dimeric structure in the asymmetric unit. Such dimers are aligned around the  $c$ -axis as if they form a helical tube with double strands, in which interdimer contacts are formed between transmembrane helices I and II of A- and B-subunits with anti-parallel orientations. **c**, The dimer interface was maintained by salt bridges between Arg 24 and Glu 25 of both monomers. **d**, Side view of the B-subunit. Helices are colour-coded from blue (helix I) to red (helix IX). Retinal (magenta spacefill) is bound to Lys 305 via a protonated Schiff base (its nitrogen atom in blue). Cys 337 is palmitoylated (grey stick representation). One phospholipid and a fatty acid in interprotein space are shown in stick representation (grey). **e**, Side view of two adjacent B-subunits related by a crystallographic two-fold axis (vertical line). Helices are drawn in cylinder representation and colour-coded according to the B-factor of C $\alpha$  atom, from blue (lowest B factor: 33 Å<sup>2</sup>) to red (highest B factor: 118 Å<sup>2</sup>). All images were drawn with Pymol<sup>32</sup>.



**Figure 3 | Structural comparisons between squid rhodopsin and bovine rhodopsin.** Squid rhodopsin is denoted by brown and magenta; bovine rhodopsin by cyan and blue.



medium is invisible in bovine rhodopsin<sup>14–16</sup>. The unique architecture of the cytoplasmic domain of squid rhodopsin provides an important insight into the recognition mechanism of a particular type of G protein.

In Fig. 3, the structure of squid rhodopsin is compared with that of bovine rhodopsin in a tetragonal  $P4_1$  crystal<sup>15</sup>. The most notable difference is seen in the third cytoplasmic (C3) loop. This difference is attributable to the extra sequence that squid rhodopsin possesses in this loop region. The multiple sequence alignment of squid and bovine rhodopsins is shown in Supplementary Fig. 1. As the C3 loop of bovine rhodopsin takes a different conformation in a different crystal form<sup>16</sup>, it has been argued that under physiological conditions the G-protein-binding site in the dark state of bovine rhodopsin is dynamically disordered. A flexible conformation of the C3 loop has been suggested previously in the  $\beta_2$ -adrenergic receptor<sup>17,18</sup>, which is another GPCR with known structure. Conversely, in squid rhodopsin the hydrophilic extension of helix VI takes a rigid conformation. As the amino acid sequence in this region is well conserved among invertebrate rhodopsins and other  $G_q$ -coupled receptors<sup>19</sup>, the extension of helices V and VI into the cytoplasmic medium can be regarded as an important structural motif for specifying the coupling mode with  $G_q$ -type G proteins.

Another unique feature of squid rhodopsin is a short  $3_{10}$  helix formed in an interhelical loop between helices VIII and IX. This  $3_{10}$  helix is dipped in the hydrophobic membrane region, whereas helix VIII is adhered to the hydrophilic surface of the lipid bilayer. Short  $3_{10}$  helices are also found in the N-terminal polypeptide and the third extracellular (E3) loop. These and other loops exposed to the extracellular medium are characterized by low temperature factors (Fig. 2e). The N-terminal polypeptide folds so as to interact with all of the extracellular loops. It may function as a scaffold for the folding of seven transmembrane helices, as proposed previously<sup>20</sup>.

### Structure of the retinal-binding pocket

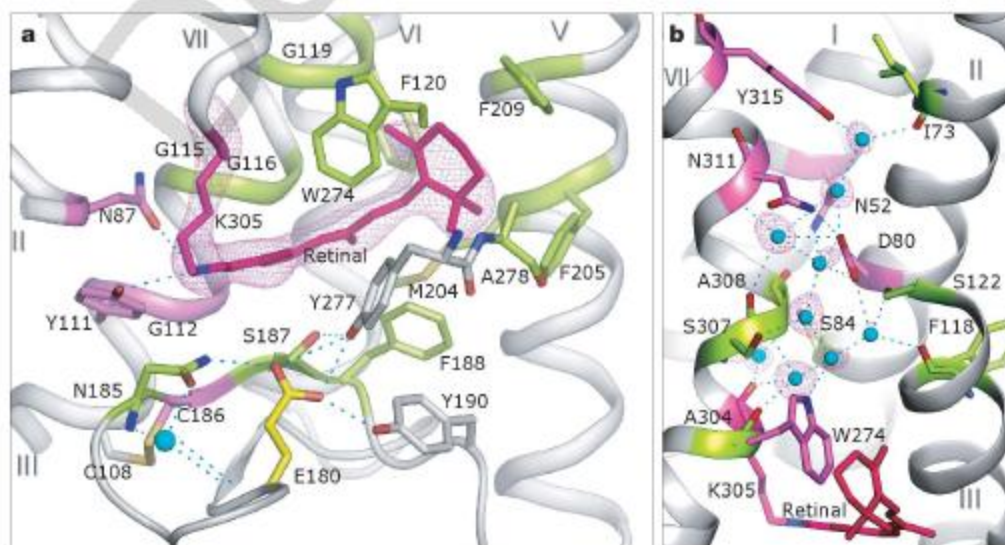
As found in most GPCRs, squid rhodopsin possesses a disulphide bond between Cys 186 in the second extracellular (E2) loop and Cys 108 in helix III (Fig. 4a). This bond is important in the proper folding of the E2 loop into the membrane region, where it forms a  $\beta$ -sheet structure that functions as a plug of the retinal-binding pocket. The 11-*cis* retinal is bound covalently to Lys 305 on helix VII. The retinal-lysine chain takes a U-shaped configuration around the indole ring of Trp 274 on helix VI. The polyene chain is orientated perpendicularly to the membrane normal, whereas the  $\beta$ -ionone ring

is rotated by  $\sim 90^\circ$  with reference to the plane of the polyene chain. Five aromatic residues (Phe 120, Phe 188, Phe 205, Phe 209 and Trp 274) are positioned around the  $\beta$ -ionone ring, and together with Met 204, they form a rigid hydrophobic pocket. The central part (C9 to C14 atoms) of retinal is sandwiched by the main chain of helix III (Gly 112, Gly 115, Gly 116 and Gly 119) and a  $\beta$ -strand (Cys 186, Ser 187 and Phe 188) in the E2 loop; its cytoplasmic side is capped by Trp 274. The present results show that residues in contact with retinal are largely altered from those seen in bovine rhodopsin. Reflecting this alteration, the retinal polyene chain takes a less distorted configuration in squid rhodopsin than in bovine rhodopsin.

The anionic residue Glu 180, which is highly conserved among visual pigments, is too far from the retinyliden-bound nitrogen to have a direct interaction (Supplementary Table 2), and the side chain of Asn 185 is located between them. It is possible that after the photoisomerization of retinal, Asn 185 moves to mediate an indirect interaction between Glu 180 and the retinal Schiff base. The hydrogen-bonding partner of the Schiff base in the dark state is either the side-chain carbonyl of Asn 87 or the OH group of Tyr 111. These residues are conserved among invertebrate rhodopsins but are replaced by glycine (Gly 89) and glutamate (Glu 113), respectively, in bovine rhodopsin. A previous spectroscopic study of octopus rhodopsin has shown that the counterpart of Tyr 111 is kept neutral in both the dark state and the acid metarhodopsin state<sup>21</sup>. This implies that in invertebrate rhodopsins, Glu 180 is the sole anionic residue existing near the Schiff base.

### Water cluster in an interhelical cavity

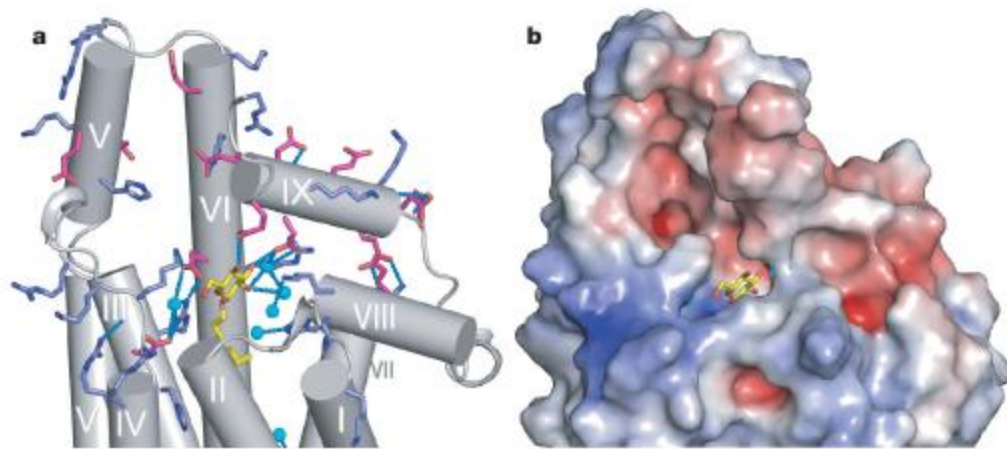
Squid rhodopsin possesses a large interhelical cavity that is filled by nine water molecules. Together with two hydrophilic residues (Asp 80 and Asn 311) existing in the middle of this cavity, these water molecules form a long hydrogen-bonding network which extends from the retinal-binding pocket to a crevice of the cytoplasmic surface (Fig. 4b). In the innermost region of this network, water molecules interact with the main-chain carbonyls of Lys 305 and Ala 304 and the indole nitrogen atom of Trp 274. The other end of the network is terminated by the phenol group of Tyr 315 in the 'NPxxY' motif which is highly conserved in GPCRs. It is noteworthy that the number of water molecules existing in the interhelical cavity is much larger than that observed in bovine rhodopsin<sup>14,15</sup>. Squid rhodopsin can therefore be classified as a typical example of an 'inside-out protein'. This classification is useful for the better understanding of the photoactivation mechanism of squid rhodopsin. A recent Fourier



**Figure 4 | Retinal-binding pocket and interhelical cavity.** **a**, Structure of the retinal-binding pocket, with an omit map of the retinal Lys 305 chain contoured at  $5\sigma$  (purple mesh). Residues within 4 Å of the retinal polyene chain are drawn in green; residues in the vicinity of the Schiff base nitrogen atom are drawn in violet. Oxygen, nitrogen and sulphur atoms are in red,

blue and gold, respectively. **b**, Interhelical cavity filled by nine water molecules (cyan), with an omit map of the water molecules contoured at  $3.8\sigma$  (purple mesh). Residues that are well conserved among GPCRs are drawn in violet. For derivation of the omit maps, diffraction intensities from a partially twinned crystal were detwinned using Detwin in CCP4<sup>33</sup>.





**Figure 5 | Cytoplasmic view of C-terminally truncated squid rhodopsin.** **a**, Octylglucoside bound to a crevice in the cytoplasmic protein surface is shown in yellow. Positively and negatively charged residues are drawn in

light blue and red, respectively. **b**, Electrostatic potential (red, negative; blue, positive) on the protein surface, viewed from the same direction as in **a**, is calculated with APBS<sup>34</sup>.

transform infrared spectroscopy study has shown that more than eight water molecules change their vibration frequencies on formation of bathorhodopsin<sup>22</sup>. It is thus suggested that a light-induced conformational change propagates towards the cytoplasmic side along the water cluster. It is conceivable that rearrangement of the water molecules inside the protein evokes a significant change in the local environment of functionally important residues such as Asp 80 and Asn 311. As these residues and their surrounding residues (Asn 52, Leu 76, Ala 77, Pro 312 and Tyr 315) are highly conserved among GPCRs, it is possible that the conformational change in the interhelical water cluster is crucial for the activation of G proteins.

#### Insights into the photoactivation mechanism

The activation mechanism of transducin,  $G_t$ -type G protein, in vertebrate photoreceptors has been extensively studied. According to a recently proposed two-state interaction model<sup>23</sup>, the GDP/GTP exchange reaction in  $G_t$  proceeds as follows: (1) the state called metarhodopsin Ib binds GDP-bound  $G_t$ ; (2) the formation of metarhodopsin II is accompanied by a large structural change in the rhodopsin- $G_t$  complex; (3) GDP is released from the complex; and (4) GTP binds to the nucleotide-free  $G_t$ . It remains unclear whether the same reaction scheme is applicable to  $G_q$ -type G proteins. The crystal structure of squid rhodopsin shows that the carboxyl group of Asp 80 interacts directly with the side chain of Asn 311. As the linkage between their equivalent residues in GPCR has been reported to be coupled with the activation of G proteins<sup>24</sup>, it is possible that the crystal structure shown here corresponds to a state in which rhodopsin is able to bind G proteins. However, as the retinal in the crystal takes an 11-*cis* configuration, an alternative explanation is that squid rhodopsin in the dark state already has the ability to form a complex with GDP-bound  $G_q$ .

In this context, it is observed that the detergent octylglucoside is bound to a crevice in the cytoplasmic surface (Fig. 5a). This crevice is surrounded by helices II, III, IV and VII and is separated by a few hydrophobic residues from the interhelical water cluster. The tail moiety of octylglucoside interacts with the hydrophobic wall of the crevice, whereas its head group interacts with several polar residues which include: Asp 132 and Arg 133 in the 'DRY' motif; Ser 64 in helix I; Lys 321 in helix VIII; and Asp 347 and Glu 351 in helix IX. Octylglucoside binding seems to stabilize the conformation of the cytoplasmic domain, as described above. It is possible that octylglucoside mimics the C-terminal polypeptide of  $G_{\alpha_q}$ , the 11 terminal residues of which are relevant to the selective recognition of a special type of GPCR<sup>25</sup>. It is notable that negatively and positively charged residues are distributed asymmetrically around a postulated binding site of the C-terminal polypeptide of  $G_{\alpha_q}$ . The calculated map of electrostatic potential on the protein surface shows that positively

charged residues (Lys 145, Lys 146, Arg 150 and Arg 151) are crowded into the N-terminal end of helix IV and the C2 loop, whereas negatively charged residues are clustered along helix IX (Fig. 5b). Such charge distribution would have a significant effect on the binding mode of the G protein. It may also affect rhodopsin dimerization, which has been shown recently to be coupled with the activation process of G proteins<sup>12</sup>.

#### Detection of the polarization plane

It has long been recognized that the invertebrate eye detects the polarization plane of visible light<sup>26,27</sup>. The crystal structure (Fig. 2) shows that two intermembrane dimers, around a crystallographic two-fold axis, form a tetrameric structure in which four retinylidene chromophores are orientated nearly parallel with one another. Under physiological conditions this tetramer may be characterized by a higher symmetry. It is possible that such tetramers are arranged in the apposed microvillar membranes so that the absorption dipole moments of all the retinal chromophores are aligned in parallel with the microvillar axis. As the microvillar membranes in each photoreceptor cell are orientated in a particular direction<sup>28</sup>, our crystal data provide an insight into the detection mechanism of the polarization plane.

Unlike vertebrate rhodopsin, the retinal chromophore in squid rhodopsin remains bound to the protein moiety under prolonged illumination. Its acid metarhodopsin state is thermally stable and, on absorption of visible light, it is hit back into the initial dark state<sup>29</sup>. At a moderate light intensity, the photoreactions of squid rhodopsin can be described by a two-photon cycle whose turnover rate is proportional approximately to  $\cos^4 \alpha$ , where  $\alpha$  is the angle that is made by the absorption dipole moment of retinal and the polarization plane of incident light. This implies that a very sensitive detection system of the polarization plane would be acquired if the two-photon cycle is involved in the activation of G protein. Further investigation into the structure–function relationship of squid rhodopsin is needed to address questions such as which type of physical quantity (light intensity, colour or polarization) does the squid eye use to perceive objects, and how do G proteins interact with the dark state of squid rhodopsin. The latter is also crucial for elucidation of the activation mechanism of human melanopsin<sup>30</sup> and other  $G_q$ -type GPCRs.

#### METHODS SUMMARY

Microvillar membranes from the photoreceptor rhabdoms in squid retina were purified using the sucrose flotation method. The purified microvillar membranes were treated with the enzyme V8-protease at 277 K for 12 h to cleave a specific site in the C-terminal extension of rhodopsin. C-terminally truncated rhodopsin was extracted selectively with octylglucoside in the presence of 30 mM zinc acetate. The final concentration of rhodopsin was  $\sim 10 \text{ mg ml}^{-1}$ . Crystals of the truncated rhodopsin were grown at 277 K by the sitting-drop



vapour-diffusion method, from a mixture solution containing ammonium sulphate and EDTA, pH 6.4. A single crystal was cryoprotected in 20% sucrose and was flash-frozen with liquid propane. X-ray diffraction measurements were performed at SPring8-BL38B1, where a frozen crystal kept at 100 K was exposed to a monochromatic X-ray beam at a wavelength of 1.0 Å with an X-ray flux rate of  $4 \times 10^{12}$  photons  $\text{mm}^{-2} \text{s}^{-1}$ .

Although the total X-ray dose per one data set ( $4 \times 10^{15}$  photons  $\text{mm}^{-2}$ ) exceeded a safety level that had been deduced from X-ray-induced absorption changes in bacteriorhodopsin crystals<sup>31</sup>, there was no detectable difference between the electron-density maps derived from two data sets taken sequentially. Therefore, no attempt was made to correct possible X-ray damage in the retinal configuration. The structure of squid rhodopsin was solved by molecular replacement, using a membrane-embedded part of bovine rhodopsin (Protein Data Bank accession 1gzm) as an initial search model. Data collection and refinement statistics are shown in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 8 January; accepted 18 March 2008.**

- Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Rev. Mol. Cell Biol.* **9**, 60–71 (2008).
- Hubbard, R. & St George, R. C. The rhodopsin system of the squid. *J. Gen. Physiol.* **41**, 501–528 (1958).
- Yarfitz, S. & Hurley, J. B. Transduction mechanisms of vertebrate and invertebrate photoreceptors. *J. Biol. Chem.* **269**, 14329–14332 (1994).
- Shichida, Y. & Imai, H. Visual pigment: G-protein-coupled receptor for light signals. *Cell. Mol. Life Sci.* **54**, 1299–1315 (1998).
- Terakita, A., Yamashita, T., Tachibanaki, S. & Shichida, Y. Selective activation of G-protein subtypes by vertebrate and invertebrate rhodopsins. *FEBS Lett.* **439**, 110–114 (1998).
- Hara-Nishimura, I. et al. Cloning and nucleotide sequence of cDNA for rhodopsin of the squid *Todarodes pacificus*. *FEBS Lett.* **317**, 5–11 (1993).
- Williamson, M. P. The structure and function of proline-rich regions in proteins. *Biochem. J.* **297**, 249–260 (1994).
- Ashida, A., Matsumoto, K., Ebrey, T. G. & Tsuda, M. A purified agonist-activated G-protein coupled receptor: truncated octopus acid metarhodopsin. *Zoolog. Sci.* **21**, 245–250 (2004).
- Murakami, M., Kitahara, R., Gotoh, T. & Kouyama, T. Crystallization and crystal properties of squid rhodopsin. *Acta Crystallogr. F* **63**, 475–479 (2007).
- Venien-Bryan, C. et al. Effect of the C-terminal proline repeats on ordered packing of squid rhodopsin and its mobility in membranes. *FEBS Lett.* **359**, 45–49 (1995).
- Saibil, H. & Hewat, E. Ordered transmembrane and extracellular structure in squid photoreceptor microvilli. *J. Cell Biol.* **105**, 19–28 (1987).
- Liang, Y. et al. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J. Biol. Chem.* **278**, 21655–21662 (2003).
- Davies, A. et al. Three-dimensional structure of an invertebrate rhodopsin and basis for ordered alignment in the photoreceptor membrane. *J. Mol. Biol.* **314**, 455–463 (2001).
- Palczewski, K. et al. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Okada, T. et al. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* **342**, 571–583 (2004).
- Li, J. et al. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
- Cherezov, V. et al. High-resolution crystal structure of an engineered human  $\beta_2$ -adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265 (2007).
- Rasmussen, S. G. F. et al. Crystal structure of the human  $\beta_2$  adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
- Horn, F. et al. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **26**, 275–279 (1998).
- Doi, T., Molday, R. S. & Khorana, H. G. Role of the intradiscal domain in rhodopsin assembly and function. *Proc. Natl Acad. Sci. USA* **87**, 4991–4995 (1990).
- Nakagawa, M. et al. How vertebrate and invertebrate visual pigments differ in their mechanism of photoactivation. *Proc. Natl Acad. Sci. USA* **96**, 6189–6192 (1999).
- Ota, T. et al. Structural changes in the Schiff base region of squid rhodopsin upon photoisomerization studied by low-temperature FTIR spectroscopy. *Biochemistry* **45**, 2845–2851 (2006).
- Morizumi, T., Imai, H. & Shichida, Y. Y. Direct observation of the complex formation of GDP-bound transducin with the rhodopsin intermediate having a visible absorption maximum in rod outer segment membranes. *Biochemistry* **44**, 9936–9943 (2005).
- Urizar, E. et al. An activation switch in the rhodopsin family of G protein-coupled receptors. *J. Biol. Chem.* **280**, 17135–17141 (2005).
- Hamm, H. E. et al. Site of G protein binding to rhodopsin mapped with synthetic peptides from the alpha subunit. *Science* **241**, 832–835 (1988).
- Jander, P., Daumer, K. & Waterman, T. H. Polarized light orientation by two hawaiian decapod cephalopods. *J. Comp. Physiol. A* **46**, 383–394 (1963).
- Saidel, W. M., Shashar, N., Schmolesky, M. T. & Hanlon, R. T. Discriminative responses of squid (*Loligo pealeii*) photoreceptors to polarized light. *Comp. Biochem. Physiol. A* **142**, 340–346 (2005).
- Saibil, H. R. An ordered membrane-cytoskeleton network in squid photoreceptor microvilli. *J. Mol. Biol.* **158**, 435–456 (1982).
- Naito, T., Nashima-Hayama, K., Ohtsu, K. & Kito, Y. Photoreactions of cephalopod rhodopsin. *Vision Res.* **21**, 935–941 (1981).
- Provencio, I. et al. Melanopsin: An opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
- Matsui, Y. et al. Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J. Mol. Biol.* **324**, 469–481 (2002).
- DeLano, W. L. The PyMOL Molecular Graphics System. (<http://www.pymol.org>) (2002).
- Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Baker, N. A. et al. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by Grant-in-Aids from the Ministry of Education, Science and Culture of Japan and partly by the National Project on Protein Structural and Functional Analyses.

**Author Contributions** T.K. and M.M. designed the project. M.M. performed all experiments. T.K. assisted in data collection and structure determination. T.K. and M.M. jointly wrote the manuscript.

**Author Information** The coordinates have been deposited in the Protein Data Bank under accession number 2z73. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.K. ([kouyama@bio.phys.nagoya-u.ac.jp](mailto:kouyama@bio.phys.nagoya-u.ac.jp)).



## METHODS

**Materials.** Fresh squid (*Todarodes pacificus*) were purchased from Yanagibashi fish market in Nagoya, Japan. All chemicals were obtained from Wako Pure Chemical Industries with the exception of the endoprotease Glu-C from *Staphylococcus aureus* V8 (V8-protease), which was purchased from Sigma-Aldrich.

**Membrane purification.** All manipulations were performed under dim red light (>640 nm) at 277 K unless otherwise stated. Microvillar membranes from rhabdomeric photoreceptor cells of squid retina were purified according to the method described previously<sup>35</sup>. Briefly, crude microvillar membranes were isolated in phosphate buffer (50 mM NaPO<sub>4</sub>, pH 6.8, 4% NaCl, 1 mM EDTA, 0.1 mM phenylmethylsulfonyl fluoride, 1 mM dithiothreitol; 0.2 µg ml<sup>-1</sup> leupeptin, 0.01% NaN<sub>3</sub>) using the sucrose flotation method. After sucrose step-gradient centrifugation of the microvillar membranes, rhodopsin-rich membranes were collected from the interface between the 20% and 32% (w/v) sucrose layers and washed with water at least three times to remove peripheral components.

**Truncation of the C-terminal extension of rhodopsin.** The purified microvillar membranes (~5 mg ml<sup>-1</sup> rhodopsin) were treated with the enzyme V8-protease (100:1 (w/w) rhodopsin:enzyme) at 277 K over 12 h to cleave at a specific site in the C-terminal extension of rhodopsin. The reaction was terminated by repeated washing of the microvillar membranes. The washed microvillar membranes were suspended in a minimum volume of water and the rhodopsin concentration was determined from the absorbance at 480 nm using the reported absorption coefficient of 35,000 mol<sup>-1</sup> cm<sup>-1</sup> (ref. 36). The membrane suspension containing ~10 mg ml<sup>-1</sup> C-terminally truncated rhodopsin was stored at 193 K until further purification.

**Selective extraction of rhodopsin.** Selective extraction of C-terminally truncated rhodopsin from the microvillar membranes was performed according to the method developed for purification of bovine rhodopsin<sup>37</sup> with minor modifications. C-terminally truncated squid rhodopsin was extracted selectively from the microvillar membranes by mixing the membrane suspension with a buffer/detergent solution giving the final concentrations: 30 mM MES, pH 6.4, 30 mM zinc acetate, and ~100 mM octylglucoside with a detergent/protein ratio of ~3.7 (w/w). After incubation at 277 K overnight, the mixture was centrifuged at 30,000 r.p.m. for 1 h and the supernatant was used within one day for crystallization.

**Crystallization.** The extracted rhodopsin was crystallized at 277 K by the sitting-drop vapour-diffusion method using commercially available crystallization kits (Vapour Diffusion Crystalplate, ICN Biomedicals). The protein solution was mixed with an equal volume of a reservoir solution containing 3.2 M ammonium sulphate, 32 mM MES pH 6.4, 38 mM EDTA and 10 mM β-mercaptoethanol. A drop (10 µl) of the mixture was equilibrated against 0.5 ml of the reservoir solution. After incubation at 277 K for >six months, hexagonal crystals had grown in size to 200 µm.

**Data collection.** A single crystal was transferred into a solution containing 30 mM octylglucoside, 3.2 M ammonium sulphate, 40 mM MES pH 6.4 and 20% sucrose. After soaking for 10 min, the crystal was flash-frozen in liquid propane held at its melting temperature and was stored in liquid nitrogen. X-ray diffraction measurements were performed at SPring8-BL38B1, where a frozen crystal kept at 100 K was exposed to a monochromatic X-ray beam at a wavelength of 1.0 Å with an X-ray flux rate of  $4 \times 10^{12}$  photons mm<sup>-2</sup> s<sup>-1</sup>. Indexing and integration of diffraction spots were carried out with Mosflm 6.2 (ref. 38). The scaling of data was carried out using SCALA in the CCP4 program suites<sup>39</sup>. Data statistics are shown in Supplementary Table 1.

**Structure refinement.** Structure refinement was done with CNS-1.1<sup>39</sup> and XtalView-4.0<sup>40</sup>. Diffraction data from the hexagonal crystal were fitted well by unit-cell parameters  $a = b = 122.6$  Å,  $c = 158.7$  Å,  $\alpha = \beta = 90^\circ$ , and  $\gamma = 120^\circ$ . Inspection of diffraction spots along the  $c$ -axis indicated the presence of a crystallographic three-fold screw axis. After various trials, the space group was assigned to  $P6_3$ . The model building was performed by the molecular-replacement method with a membrane-embedded part of bovine rhodopsin (Protein Data Bank accession 1GZM) as an initial search model. Diffraction intensities from a partially twinned crystal (twinning fraction ~0.24) were detwinned by CCP4 and the resultant intensities  $I_{\text{detwin}}(hkl)$  were used to build a structural model at 2.6 Å resolution. This model was further refined with a higher resolution data set from a nearly perfect twin. In this case, diffraction intensities were detwinned by the equation:

$$I_{\text{detwin}}(hkl) = \{2 \langle I_{\text{obs}}(hkl) \rangle + I_{\text{calc}}(hkl) - I_{\text{calc}}(k, h, -l)\} / 2$$

where  $\langle I_{\text{obs}}(hkl) \rangle = (I_{\text{obs}}(hkl) + I_{\text{obs}}(k, h, -l)) / 2$  and  $I_{\text{calc}}$  is the intensity calculated from the model.

35. Kito, Y., Naito, T. & Nashima, K. Purification of squid and octopus rhodopsin. *Methods Enzymol.* **81**, 167–171 (1982).
36. Suzuki, T., Uji, K. & Kito, Y. Studies on cephalopod rhodopsin: photoisomerization of the chromophore. *Biochim. Biophys. Acta* **428**, 321–338 (1976).
37. Okada, T., Takeda, K. & Kouyama, T. Highly selective separation of rhodopsin from bovine rod outer segment membranes using combination of divalent cation and alkyl (thio) glucoside. *Photochem. Photobiol.* **67**, 495–499 (1998).
38. Steller, I., Bolotovsky, R. & Rossmann, M. G. An algorithm for automatic indexing of oscillation images using Fourier analysis. *J. Appl. Crystallogr.* **30**, 1036–1040 (1997).
39. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
40. McRee, D. E. *Practical Protein Crystallography* (Academic, San Diego, 1993).



## ARTICLES

# Nucleosome organization in the *Drosophila* genome

Travis N. Mavrich<sup>1,2\*</sup>, Cizhong Jiang<sup>1,2\*</sup>, Ilya P. Ioshikhes<sup>3</sup>, Xiaoyong Li<sup>4</sup>, Bryan J. Venters<sup>1,2</sup>, Sara J. Zanton<sup>1,2</sup>, Lynn P. Tomsho<sup>2</sup>, Ji Qi<sup>2</sup>, Robert L. Glaser<sup>5</sup>, Stephan C. Schuster<sup>2</sup>, David S. Gilmour<sup>1</sup>, Istvan Albert<sup>2</sup> & B. Franklin Pugh<sup>1,2</sup>

**Comparative genomics of nucleosome positions provides a powerful means for understanding how the organization of chromatin and the transcription machinery co-evolve. Here we produce a high-resolution reference map of H2A.Z and bulk nucleosome locations across the genome of the fly *Drosophila melanogaster* and compare it to that from the yeast *Saccharomyces cerevisiae*. Like *Saccharomyces*, *Drosophila* nucleosomes are organized around active transcription start sites in a canonical  $-1$ , nucleosome-free region,  $+1$  arrangement. However, *Drosophila* does not incorporate H2A.Z into the  $-1$  nucleosome and does not bury its transcriptional start site in the  $+1$  nucleosome. At thousands of genes, RNA polymerase II engages the  $+1$  nucleosome and pauses. How the transcription initiation machinery contends with the  $+1$  nucleosome seems to be fundamentally different across major eukaryotic lines.**

Knowledge of the precise location of nucleosomes in a genome is essential to understand the context in which chromosomal processes such as transcription and DNA replication operate. A common theme to emerge from recent genome-wide maps of nucleosome locations is a general deficiency of nucleosomes in promoter regions and an enrichment of certain histone modifications towards the 5' end of genes<sup>1–7</sup>. A high resolution genomic map of nucleosome locations in the budding yeast *S. cerevisiae* has further revealed the nucleosomal context of *cis*-regulatory elements and transcriptional start sites<sup>1–7</sup>. However, such context has not been established in multicellular eukaryotes, and so fundamental questions remain: Is there a common theme by which genes of multicellular eukaryotes position their nucleosomes with respect to functional chromosomal elements? Are such themes and their underlying rules evolutionarily conserved across eukaryotes? What are the functional implications for those themes that differ across the major eukaryotic lines? To address these questions, we have produced a genome-wide high-resolution map of H2A.Z (also known as H2Av) and bulk nucleosome locations in the embryo of the fruitfly *D. melanogaster*. H2A.Z is widely distributed in *Drosophila*<sup>8</sup>, but some evidence points to specialized roles<sup>9,10</sup>. In *Saccharomyces*, H2A.Z replaces H2A at the 5' end of active genes<sup>11–14</sup>, and thus provides a focused representation of promoter chromatin architecture.

*Drosophila* embryos are composed of a wide variety of cell types in which subsets of genes may elicit distinct gene expression programmes<sup>15,16</sup>. Global gene expression profiles during all stages of *Drosophila* development from 8–12 h post fertilization to a young adult fly are correlated (Supplementary Fig. 1), which possibly reflects the broad expression pattern of the large repertoire of house-keeping genes in most cell types during development<sup>15,16</sup>. This general spatial and temporal independence of gene expression provides impetus to use whole embryos to develop a reference nucleosome map. Indeed, our map reveals that nucleosomes are generally well organized, despite cell type heterogeneity.

## 'Closed' and 'open' chromatin organization

Embryos were treated with formaldehyde, and H2A.Z nucleosome core particles were immunopurified (Supplementary Figs 2 and 3). H2A.Z-containing nucleosomes (652,738) were sequenced (Supplementary Fig. 4) and mapped to 207,025 consensus locations in the *Drosophila* r5.2 reference genome (Fig. 1a and Supplementary Fig. 2b, see browser at <http://atlas.bx.psu.edu/>), thereby providing >3-fold depth of coverage (Supplementary Fig. 5). Correction for micrococcal nuclease (MNase) digestion bias was imposed (Supplementary Fig. 6). Those 112,750 nucleosomes detected three or more times were further analysed, although patterns were identical when all nucleosomes were analysed. The internal median error of the data was 4 bp (Supplementary Fig. 7).

Figure 1b displays the predominant embryonic distribution of H2A.Z nucleosomes relative to the transcription start site (TSS) of all coding genes, and is compared to the pattern previously derived from *Saccharomyces*<sup>1</sup>. Patterns around noncoding genes are shown in Supplementary Fig. 8. Eighty five per cent of *Drosophila* coding genes (11,994 of the 14,143) contained at least one H2A.Z nucleosome (detected three or more times) within 1 kb of the TSS. H2A.Z levels correlated with gene expression (Fig. 1c and Supplementary Fig. 9), as has been seen on individual genes and in *Saccharomyces*<sup>12,13,17</sup>.

H2A.Z nucleosomes were predominantly distributed at 175-bp intervals from the TSS (compared to 165-bp in *Saccharomyces*<sup>1</sup>, Fig. 1b), demonstrating that a predominant organizational pattern exists for H2A.Z nucleosomes in *Drosophila* embryos that transcends a spatial and temporal context. The H2A.Z pattern was compared to the distribution of bulk nucleosomes (that is, those containing any combination of H2A.Z and H2A), determined using high-density tiling arrays (36-bp probe spacing). Within genic regions, the same organizational pattern was found (Supplementary Fig. 10). For both data sets, a nucleosome-depleted region was evident immediately upstream of the  $+1$  nucleosome, which probably reflects a nucleosome-free core promoter region (NFR), as first detected in

<sup>1</sup>Center for Gene Regulation, and <sup>2</sup>Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. <sup>3</sup>Department of Biomedical Informatics and Department of Molecular and Cellular Biochemistry, Davis Heart and Lung Research Institute, The Ohio State University, Columbus, Ohio 43210, USA. <sup>4</sup>Berkeley Drosophila Transcription Network Project, Genomics Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, California 94720, USA. <sup>5</sup>Wadsworth Center, New York State Department of Health and Department of Biomedical Sciences, State University of New York, Albany, New York 12201-2002, USA.

\*These authors contributed equally to this work.



*Saccharomyces*<sup>2</sup>. Like *Saccharomyces*, a -1 nucleosome was detected ~180 bp upstream of the TSS. However, in contrast, it lacked H2A.Z.

Surprisingly, the genic array of *Drosophila* nucleosomes started ~75 bp further downstream from the equivalent position in *Saccharomyces*, placing the +1 nucleosome at +135 (Fig. 1b and Supplementary Fig. 10). This shift has important implications in how the TSS is presented to RNA polymerase II (Pol II). In *Saccharomyces*, the TSS resides within the nucleosome border, potentially allowing the nucleosome to regulate start-site selection and efficiency<sup>1</sup>. In *Drosophila*, the predominant arrangement of nucleosomes might allow unimpeded access to the TSS, with potential blockage occurring downstream after initiation.

*Drosophila* have well-defined core promoter elements, such as TATA, initiator (Inr), downstream promoter element (DPE) and motif ten element (MTE), which bind to the general transcription machinery<sup>18–22</sup>, although these elements are not found in most genes. For genes lacking these core promoter elements or having a DPE, the canonical nucleosome organization was observed (black pattern in Supplementary Fig. 11), which was more robust when only H2A.Z-containing nucleosomes were examined (blue pattern). In contrast, genes containing TATA, Inr or MTE had a diminished canonical nucleosome organization and a diminished NFR, indicating that these classes of genes may have a more compact and gene-specific chromatin architecture, including a positioned nucleosome over the TSS. Consequently, they might be more dependent on chromatin remodelling for expression. When genes become transcriptionally competent, resident nucleosomes could adopt a more open and canonical organization, which includes replacing H2A with H2A.Z. Three observations support this hypothesis. First, H2A.Z and bulk nucleosomes at highly expressed genes were more uniformly organized than those at genes with a lower expression (Supplementary Fig. 9). Second, bulk nucleosomes for genes that contained H2A.Z at their 5' end displayed the canonical pattern, whereas those lacking H2A.Z did not (Supplementary Fig. 10, black plot versus red trace). Third, within any class of genes except those having an Inr, H2A.Z nucleosomes adopted a more canonical organization than the bulk set of nucleosomes (Supplementary Fig. 11). These results suggest that transcription and the presence of H2A.Z are linked to an open and uniform chromatin architecture at promoter regions.

### DNA motif organization around nucleosomes

Recent genome sequencing of 12 *Drosophila* species of differing evolutionary distance has provided an unprecedented opportunity to

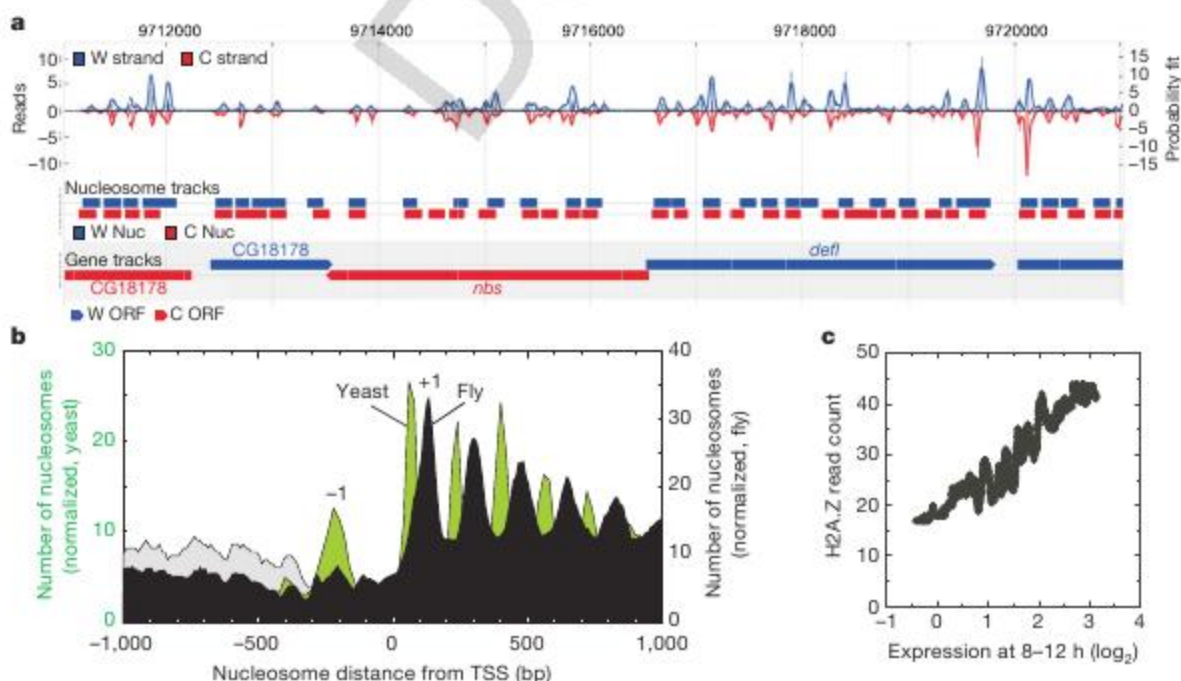
identify conserved DNA sequence motifs<sup>23</sup>. By comparing the distribution of motifs around the TSS<sup>23</sup>, we found four recurring patterns: 27 motifs were classified as 'nucleosomal', 57 as 'anti-nucleosomal', 12 as 'fixed', and 98 as 'random' (left panels in Fig. 2a and Supplementary Fig. 12). Nucleosomal and anti-nucleosomal patterns matched the general distribution of where nucleosomes were relatively enriched or depleted, respectively, relative to the TSS (see Fig. 1b). Fixed elements were at a defined distance from the TSS, and random elements lacked patterning. The nucleosomal and anti-nucleosomal patterns suggest that certain motifs are organized to be downstream of the TSS in the midst of nucleosomal arrays, whereas others are organized to be upstream of the TSS, where nucleosomes are relatively depleted.

We examined the organizational relationship of these DNA motifs to individual H2A.Z nucleosomes within the whole genome (right panels of Fig. 2a and Supplementary Fig. 12, and all motifs in Fig. 2b). Notably, nucleosomal motifs were consistently enriched on the H2A.Z nucleosome surface, whereas anti-nucleosomal motifs were consistently depleted. Individual fixed motifs were mostly depleted of H2A.Z nucleosomes. These findings along with several controls (Supplementary Fig. 13) suggest that motifs and nucleosomes adopt a preferred organization around each other, regardless of their genomic location. This organization could be linked to co-evolution of base sequence composition bias in and around nucleosomes. The functional importance of such context remains to be determined.

### Nucleosome-positioning sequences

We examined whether the positions of *Drosophila* H2A.Z nucleosomes are at least partly defined by the underlying DNA sequence pattern, and investigated whether such a pattern might be evolutionarily conserved. We determined the frequency of dinucleotides across *Drosophila* H2A.Z nucleosomal DNA because 10-bp periodic patterns of certain dinucleotides enhance the wrapping and positioning of DNA around the histone core (Fig. 3a and Supplementary Fig. 14). As seen in *Saccharomyces*, 10-bp periodic patterns of A and/or T (A/T) dinucleotides running counter-phase to G/C dinucleotides was observed. The modest amplitudes of the pattern suggest that such periodicities are infrequent, and are thus used selectively (that is, most nucleosomes lack underlying positioning signals).

We further investigated the rules of nucleosome positioning by scanning promoter regions for correlations to nucleosome positioning sequences previously identified for a relatively small number of eukaryotic nucleosomes<sup>24</sup>, in which AA/TT (yeast<sup>25</sup> and worms<sup>26</sup>) or CC/GG (human)<sup>27</sup> dinucleotides occur in a biased and/or periodic



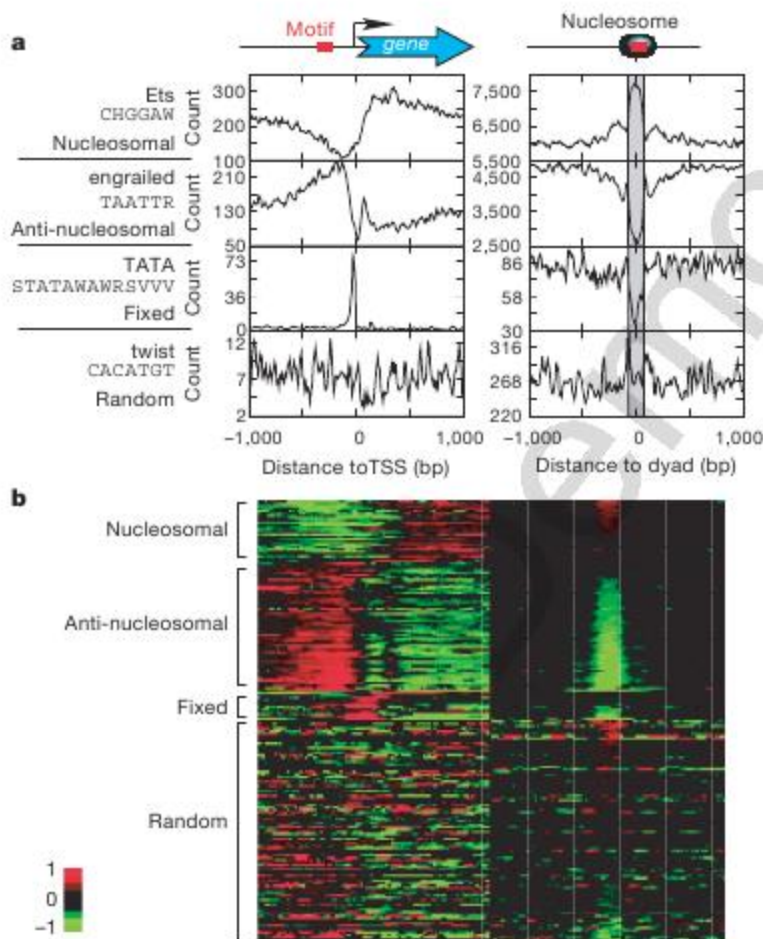
**Figure 1 | H2A.Z nucleosomal organization around the 5' end of *Drosophila* genes.** **a**, Browser shot of an arbitrary locus (*nbs-defl*). The bar graph represents the number of 'W' or 'C' (top and bottom traces, respectively) strand reads mapped to each coordinate. **b**, Composite distribution of H2A.Z nucleosomes relative to the TSS. Nearby genes were either included (grey) or eliminated (black) from the analysis, and normalized accordingly. The equivalent *Saccharomyces* profile is shown in green. **c**, Correlation of the number of H2A.Z nucleosomal sequencing reads (per gene) to messenger RNA levels in embryos (8–12 h).



arrangement across nucleosomal DNA. Unlike in yeast, the AA/TT positioning pattern failed to identify nucleosome locations (Fig. 3b, black trace). However, the CC/GG pattern (Supplementary Fig. 15) reproduced the exact position of the +1 nucleosomes (Fig. 3b, red trace), indicating that the *Drosophila* +1 nucleosome may be positioned in part by CC/GG-based positioning sequences that are used preferentially in metazoans. Consistent with this, +1 nucleosomes are highly positioned around the 5' end of genes (Supplementary Fig. 16).

### The ends of genes are nucleosome-free

Despite H2A.Z being enriched at the 5' end of genes, substantial levels were detected throughout the genome, which allowed us to examine nucleosome organization at the 3' end of genes (Fig. 4a and Supplementary Fig. 17a). Notably, H2A.Z nucleosome levels spiked near the open reading frame (ORF) end points and then dropped precipitously further downstream into the intergenic regions, where transcripts terminate. The spike occurred ~30 bp upstream from the stop codon and ~160 bp upstream of the transcript poly(A) site. A similar nucleosome drop-off was seen when bulk nucleosomes were examined (Supplementary Fig. 17b), but was not evident at genes that lacked H2A.Z. Thus, like the 5' end, the presence of H2A.Z may be linked to a more open chromatin architecture at the 3' end of genes. The change in nucleosome density coincided with alterations in nucleosome positioning sequences (Fig. 4b). Thus, such '3' NFRs' might be defined in part by the



**Figure 2 | Organization of conserved DNA motifs around TSSs (left) and nucleosomes (right).** **a**, Distribution of representative conserved DNA motifs from four distribution classes. The motif name (TATA) or the protein that binds to the motif (E-26-specific or Ets, engrailed and twist) is indicated above the motif DNA sequence **b**, Distribution of all motifs plotted with Treeview<sup>46</sup>. Bin counts from all motifs in Supplementary Fig. 12 were converted to fold deviations from the regional average ( $\pm 1$  kb), converted to a log<sub>2</sub> scale, and plotted. Red, black and green denote above, near or below average deviations, respectively.

360

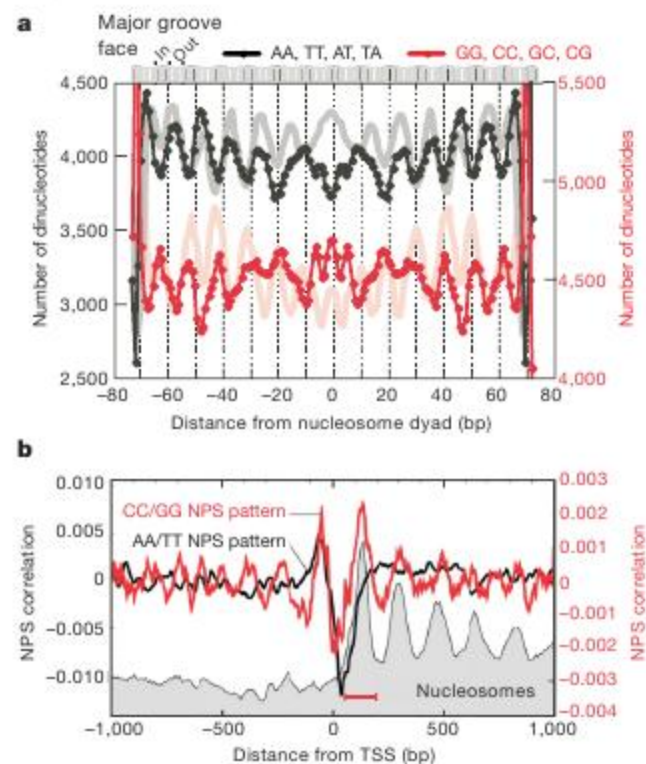
underlying DNA sequence. Conceivably, 3' NFRs might function in transcription termination.

### RNA polymerase contacts the +1 nucleosome

The location of the +1 nucleosome at the 5' end of genes is notable because its upstream border resides at approximately +62 (relative to the TSS), which is near where Pol II pauses during the transcription cycle<sup>3,28–32</sup>. To examine the potential linkage between Pol II pausing and nucleosome positions, we first determined the genome-wide location of Pol II in embryos at 1,956 putatively paused genes (Fig. 5a). Pol II was concentrated in a ~300 bp region that peaked around +90, which overlaps the region bound by the +1 nucleosome; this is consistent with other recent placements<sup>30–32</sup>. Indeed, the distribution of paused Pol II, as directly measured by permanganate reactivity of thymines on a statistically robust subset of ~50 genes (yellow trace in Fig. 5a and Supplementary Fig. 18a), indicates that pausing occurs between +20 and +50, with the centre at +35 (ref. 30). This high-resolution permanganate footprinting data, which represents the most definitive means of assessing Pol II pausing, places the front edge of Pol II (~16 bp downstream of the bubble<sup>33</sup>) within ~10 bp of the +1 nucleosome border.

The location of the +1 H2A.Z nucleosome was similar (but not identical) whether or not paused Pol II was present (Fig. 5b), indicating that Pol II was not likely to be the cause of the nucleosome shift compared to *Saccharomyces*. Instead, the positioned +1 nucleosome might be contributing to Pol II pausing, which is consistent with other studies<sup>34–37</sup>. Other factors including negative elongation factor (NELF) are likely to make significant contributions to pausing as well<sup>30,38,39</sup>.

Intriguingly, genes that contained a paused Pol II showed a ~10 bp downstream shift of H2A.Z nucleosomes ( $P$ -value =  $10^{-9}$ ; Fig. 5b). The same shift was observed if H2A.Z sequencing reads (rather than



**Figure 3 | Positioning properties of *Drosophila* nucleosomes and DNA.**

**a**, Composite distribution of WW (AA, TT, AT and TA represented by a black trace) and SS (GG, CC, GC and CG represented by the red trace) dinucleotides along the 147-bp axis of nucleosomal DNA ( $P$ -value, 0). The equivalent yeast profile<sup>1</sup> is shown in light shading in the background. **b**, Average correlation of all *Drosophila* promoter regions to nucleosome positioning sequence (NPS) patterns, comparing an AA/TT and a CC/GG pattern. The distribution of H2A.Z nucleosomes from Fig. 1b is shown as a grey backdrop. The span of the +1 nucleosome is indicated by the red horizontal bar.



nucleosomes) or bulk nucleosomes are plotted (Supplementary Fig. 19a, b). The shift suggests that, as part of the pausing process, Pol II collides with the +1 nucleosome, possibly displacing it downstream by one turn of the DNA helix. If the downstream nucleosomes are positioned mainly by the principles of statistical positioning<sup>40,41</sup>, rather than the underlying DNA sequence, then a shift of the +1 nucleosome is expected to have a ripple effect on downstream nucleosomes.

To test the prediction that Pol II is engaging the +1 nucleosome, bulk mononucleosomes were prepared from formaldehyde-crosslinked embryos and immunoprecipitated with antibodies directed against Pol II. DNA corresponding to mononucleosomes (~150 bp) was gel-purified and mapped to the entire *Drosophila* genome with high-resolution tiling arrays. Figure 5c (black trace) shows that the distribution of nucleosome–Pol II crosslinking at Pol II-paused genes peaked at the +1 nucleosome. This was not seen at genes lacking a paused Pol II or H2A.Z. The selective enrichment at +1 demonstrates that Pol II is predominantly engaged with the +1 nucleosome, and therefore the +1 nucleosome may be instrumental in establishing the paused state.

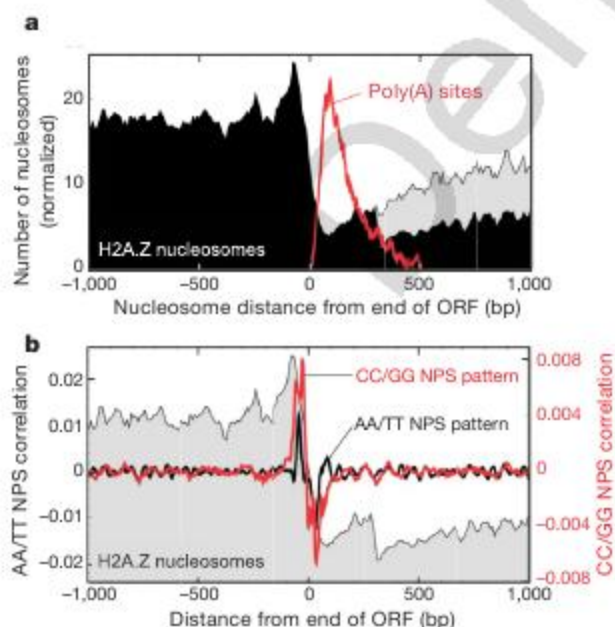
## Conclusions

The high resolution map of *Drosophila* nucleosomes reveals evolutionarily conserved and divergent principles of nucleosome organization. Genes that possess H2A.Z nucleosomes are likely to have experienced a transcription event. They tend to have nucleosome-free promoter and termination regions and intervening arrays of uniformly positioned nucleosomes that become less uniform towards the 3' end of the gene. H2A.Z nucleosomes in general might not block assembly of the transcription machinery at transcriptionally 'experienced' promoters. However, repressed promoters or those containing Inr elements do seem to have an H2A nucleosome over the TSS.

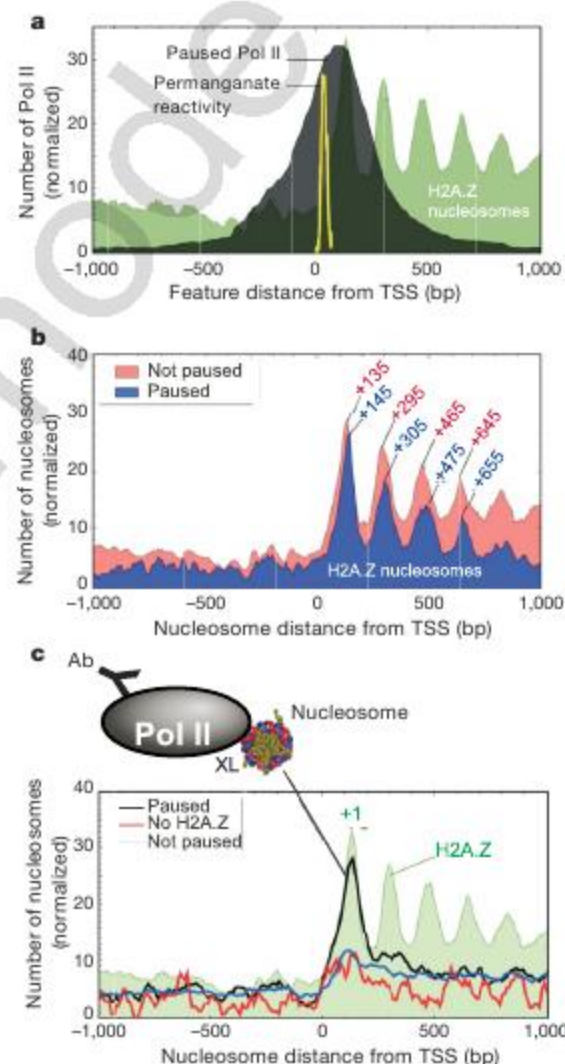
Conserved DNA sequence motifs (and thus any proteins that bind to them) tend to have an organizational relationship with nucleosomes. 'Anti-nucleosomal' motifs including those for proteins such as engrailed, even-skipped, fushi tarazu, giant, hunchback and knirps tend to be located upstream of the TSS and might contribute to the exclusion of nucleosomes over the core promoter. Indeed some

have anti-nucleosomal activity<sup>42,43</sup>. 'Nucleosomal' motifs include sites for achaete, antennapedia, dorsal, tramtrack and others. Their preference for locations downstream of the TSS where nucleosomes are well organized raises the possibility that they contribute to nucleosome organization.

In *Saccharomyces*, the location of the TSS just inside the +1 nucleosome border allows the nucleosome potentially to exert control over initiation; however, in *Drosophila*, most genes might position the +1 nucleosome to interact with a transcriptionally engaged paused polymerase. It is not known whether the +1 nucleosome is causative or just participatory in the pausing. It is now becoming clear that metazoans also regulate transcription through Pol II pausing rather than solely through transcription complex assembly<sup>3,31,32,44</sup>. The nucleosome map and its context to DNA regulatory elements, presented here, provides a framework for designing



**Figure 4 | H2A.Z nucleosomal organization around the 3' end of *Drosophila* genes.** **a**, Composite distribution of nucleosomes relative to ORF end points. Also shown is the distribution of transcript termination sites (poly(A) sites) in red. Nearby genes were either included (grey) or eliminated (black) from the analysis. **b**, Average correlation of all *Drosophila* gene terminal regions to AA/TT or CC/GG NPS patterns. The nucleosome profile from **a** is shown as a grey backdrop.



**Figure 5 | Distribution of Pol II and Pol II-engaged nucleosomes around the 5' end of genes.** **a**, Genome-wide location of 'paused' Pol II relative to the TSS. Chromatin immunoprecipitation (ChIP)–chip genomic profiling of Pol II was conducted on *Drosophila* embryos. The black filled plot shows the distribution of Pol II at 1,956 'paused' Pol II genes (see Methods). The yellow trace shows the distribution of permanganate-reactive thymines (an indicator of pausing) in 50 genes that undergo pausing<sup>30</sup>. The distribution of Pol II at these 50 genes is similar to the bulk profile (Supplementary Fig. 18a). **b**, H2A.Z nucleosomal distribution at 1,956 genes that contained 'paused' Pol II. The 'not paused' class represents those for which Pol II is either absent or not paused. **c**, Distribution of Pol II-bound nucleosomes. Pol II ChIP was performed on bulk MNase-digested mono-nucleosomal DNA and hybridized to genome-wide tiling arrays. The three traces include distributions at 1,956 Pol II 'paused' genes and control distributions either at 788 genes that have 'no H2A.Z' within 1 kb of the TSS or at 8,736 genes of the 'not paused' class.



experiments and analysing existing data to understand how metazoans regulate transcription.

## METHODS SUMMARY

*D. melanogaster* embryos (0–12 h) were collected and crosslinked with formaldehyde. H2A.Z was immunoprecipitated from chromatin digested with MNase. Mononucleosomal DNA was gel-purified and sequenced using Roche GS20/FLX pyrosequencing technology<sup>1,45</sup>. Chromatin from crosslinked embryos was also solubilized by sonication and/or MNase digestion, where indicated, and Pol II immunoprecipitated. Bulk nucleosomes were not immunoprecipitated. MNase-treated samples were gel-purified in the 75–200 bp range. DNA samples were then hybridized to Affymetrix *Drosophila* tiling microarrays (36 bp average probe spacing).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 7 February; accepted 26 March 2008.

Published online 13 April 2008.

- Albert, I. et al. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**, 572–576 (2007).
- Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
- Lee, W. et al. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39**, 1235–1244 (2007).
- Mito, Y., Henikoff, J. G. & Henikoff, S. Genome-scale profiling of histone H3.3 replacement patterns. *Nature Genet.* **37**, 1090–1097 (2005).
- Pokholok, D. K. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).
- Yuan, G. C. et al. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
- Leach, T. J. et al. Histone H2A.Z is widely but nonrandomly distributed in chromosomes of *Drosophila melanogaster*. *J. Biol. Chem.* **275**, 23267–23272 (2000).
- Updike, D. L. & Mango, S. E. Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA. *PLoS Genet.* **2**, e161 (2006).
- Swaminathan, J., Baxter, E. M. & Corces, V. G. The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of *Drosophila* heterochromatin. *Genes Dev.* **19**, 65–76 (2005).
- Lieb, J. D. & Clarke, N. D. Control of transcription through intragenic patterns of nucleosome composition. *Cell* **123**, 1187–1190 (2005).
- Raisner, R. M. et al. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**, 233–248 (2005).
- Zhang, H., Roberts, D. N. & Cairns, B. R. Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell* **123**, 219–231 (2005).
- Li, B. et al. Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl Acad. Sci. USA* **102**, 18385–18390 (2005).
- Hild, M. et al. An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5**, R3 (2003).
- Tomancak, P. et al. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8**, R145 (2007).
- Bruce, K. et al. The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. *Nucleic Acids Res.* **33**, 5633–5639 (2005).
- Purnell, B. A., Emanuel, P. A. & Gilmour, D. S. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev.* **8**, 830–842 (1994).
- Kutach, A. K. & Kadonaga, J. T. The downstream promoter element DPE appears to be as widely used as the TATA box in *Drosophila* core promoters. *Mol. Cell. Biol.* **20**, 4754–4764 (2000).
- Lim, C. Y. et al. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev.* **18**, 1606–1617 (2004).
- Biggin, M. D. & Tjian, R. Transcription factors that activate the *Ultrabithorax* promoter in developmentally staged extracts. *Cell* **53**, 699–711 (1988).
- Soeller, W. C., Oh, C. E. & Kornberg, T. B. Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. *Mol. Cell. Biol.* **13**, 7961–7970 (1993).
- Stark, A. et al. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219–232 (2007).
- Ioshikhes, I., Bolshoy, A., Derenshteyn, K., Borodovsky, M. & Trifonov, E. N. Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.* **262**, 129–139 (1996).
- Ioshikhes, I. P., Albert, I., Zanton, S. J. & Pugh, B. F. Nucleosome positions predicted through comparative genomics. *Nature Genet.* **38**, 1210–1215 (2006).
- Johnson, S. M., Tan, F. J., McCullough, H. L., Riordan, D. P. & Fire, A. Z. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* **16**, 1505–1516 (2006).
- Kogan, S. B., Kato, M., Kiyama, R. & Trifonov, E. N. Sequence structure of human nucleosome DNA. *J. Biomol. Struct. Dyn.* **24**, 43–48 (2006).
- Gilmour, D. S. & Lis, J. T. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells. *Mol. Cell. Biol.* **6**, 3984–3989 (1986).
- Law, A., Hirayoshi, K., O'Brien, T. & Lis, J. T. Direct cloning of DNA that interacts *in vivo* with a specific protein: application to RNA polymerase II and sites of pausing in *Drosophila*. *Nucleic Acids Res.* **26**, 919–924 (1998).
- Lee, C. et al. NELF and GAGA factor are linked to promoter proximal pausing at many genes in *Drosophila*. *Mol. Cell. Biol.* (in the press).
- Muse, G. W. et al. RNA polymerase is poised for activation across the genome. *Nature Genet.* **39**, 1507–1511 (2007).
- Zeitlinger, J. et al. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39**, 1512–1516 (2007).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Brown, S. A., Imbalzano, A. N. & Kingston, R. E. Activator-dependent regulation of transcriptional pausing on nucleosomal templates. *Genes Dev.* **10**, 1479–1490 (1996).
- Brown, S. A. & Kingston, R. E. Disruption of downstream chromatin directed by a transcriptional activator. *Genes Dev.* **11**, 3116–3121 (1997).
- Carey, M., Li, B. & Workman, J. L. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol. Cell* **24**, 481–487 (2006).
- Bondarenko, V. A. et al. Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol. Cell* **24**, 469–479 (2006).
- Renner, D. B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D. H. A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.* **276**, 42601–42609 (2001).
- Wu, C. H. et al. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in *Drosophila*. *Genes Dev.* **17**, 1402–1414 (2003).
- Kornberg, R. The location of nucleosomes in chromatin: specific or statistical. *Nature* **292**, 579–580 (1981).
- Kornberg, R. D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16**, 6677–6690 (1988).
- Lehmann, M. Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure. *Trends Genet.* **20**, 15–22 (2004).
- Mito, Y., Henikoff, J. G. & Henikoff, S. Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**, 1408–1411 (2007).
- Lis, J. T. Imaging *Drosophila* gene activation and polymerase pausing *in vivo*. *Nature* **450**, 198–202 (2007).
- Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by grants HG004160 (B.F.P.) and GM47477 (D.S.G.). We thank M. Biggin for early access to the Pol II chromatin immunoprecipitation (ChIP)-chip data, R. Fan for supplying the antibody raised against the *Drosophila* Pol II subunit Rpb3, and C. Lee for help in identifying paused Pol II.

**Author Contributions** T.N.M. prepared and purified the nucleosomes including Pol II-bound nucleosomes; C.J. analysed the nucleosome-mapping data and its relationship to other genomic features; I.P.I. performed computational analyses related to nucleosome-positioning sequences; X.L. conducted ChIP-chip on Pol II; B.J.V. conducted ChIP-chip and analysis on GAF; S.J.Z. provided bioinformatics support; L.P.T. constructed libraries and sequenced nucleosomal DNA; J.Q. mapped sequencing reads to the yeast genome; R.L.G. provided H2A.Z antibodies; S.C.S. directed the DNA-sequencing phase; D.S.G. directed embryo preparations and helped to interpret the data; I.A. developed computational approaches to derive nucleosome maps from the read locations and developed the associated browser; and B.F.P. directed the project, interpreted the data and wrote the paper.

**Author Information** Sequence data are deposited in the NCBI Trace Archives T1 SRA000283 under the Sequencing Center designation 'CCGB', and microarray data are deposited in the ArrayExpress under accession numbers E-MEXP-1515, E-MEXP-1519 and E-MEXP-1520. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.F.P. (bfp2@psu.edu).



## METHODS

**Egg collection, dechoriation, crosslink and nuclei preparation.** Eight grams of Oregon R embryos 0–12 h old were collected and crosslinked at a time, similarly to what has been described previously<sup>47,48</sup>. The embryos were dechorionated for 90 s, washed, and equally divided into eight 50 ml tubes containing 10 ml ChIP-FIX (50 mM HEPES, pH 7.6, 100 mM NaCl, 0.1 mM EDTA, 0.5 mM EGTA, 2% formaldehyde) and 30 ml heptane. The tubes were vigorously shaken for 15 min at 25 °C and then centrifuged at 1,500g for 1 min at 25 °C, followed by the removal of the aqueous layer. The embryos were washed as follows: once with PBS + 0.01% TritonX-100 + 0.125 M glycine, and twice with PBS + 0.01% TritonX-100. For each wash, 10 ml of the indicated buffer was added, the tubes were shaken for 1 min and centrifuged as before, and the aqueous layer was removed (except after the last wash, in which the heptane was also removed). The embryos were washed once more with 20 ml PBS + 0.01% TritonX-100.

The crosslinked embryos were resuspended in 40 ml homogenization buffer (10 mM HEPES, pH 7.6, 0.3 M sucrose, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.5 mM EGTA, 1 mM DTT, 1 mM sodium bisulphite, 0.2 mM PMSF) in a dounce with 10 strokes of the loose pestle and 15 strokes of the tight pestle, while on ice. The sample was centrifuged at 2,000g for 10 min at 4 °C, the supernatant was removed, and the pellet was washed once in 10 ml NPS buffer. After centrifuging at 16,000g for 5 min at 4 °C, the supernatant was removed and the pellet was frozen in liquid nitrogen and stored at –80 °C until ready for chromatin digestion.

**H2A.Z nucleosome preparations.** See Supplementary Fig. 2 for a schematic of the procedure. MNase digestion of chromatin was carried out similarly to what has been described previously<sup>1</sup>. Fifteen grams of embryos were thawed and resuspended, using a dounce, in a total volume of 36 ml in NPS buffer + 1 mM  $\beta$ -mercaptoethanol. MNase (40 kU) was added to the sample and incubated for 2 h at 25 °C on a rotator. The sample was chilled on ice for 10 min and EDTA was added to 10 mM final concentration to quench the digestion. The sample was then centrifuged at 15,000g for 10 min at 4 °C and the supernatant was discarded. The pellet was washed twice with 36 ml formaldehyde (FA) lysis buffer + SDS and centrifuged as before, discarding the supernatant each time.

To solubilize digested chromatin, the sample was briefly sonicated. The sample was resuspended in 18 ml FA lysis buffer + SDS, and equally divided into fifteen 15-ml tubes. Two samples were sonicated simultaneously in a Diagenode Bioruptor on medium power for five sessions (each session consisting of 30 s ON and 30 s OFF). Samples were then transferred to 1.7-ml tubes and centrifuged at 16,000g for 10 min at 4 °C. The pellets were discarded and the supernatants were frozen at –20 °C until ready for ChIP.

The supernatants were thawed on ice and centrifuged at 15,000g for 10 min at 4 °C to remove any debris. Solubilized digested chromatin from 15 g of embryos was combined and the volume was increased fourfold with FA lysis buffer to dilute out the SDS. After this, the chromatin was divided equally into five 15-ml tubes. To each tube, 170  $\mu$ l anti-H2A.Z antibody<sup>4</sup> was added and samples were incubated for 14 h at 4 °C on a rotator. Chromatin was pre-cleared by incubating with 115  $\mu$ l bed volume Sepharose 4B (Amersham, 17-0120-01) for 15 min at 4 °C on a rotator. The resin was centrifuged at 2,000g for 3 min at 25 °C, and the supernatants were transferred to a new tube containing 115  $\mu$ l bed volume protein A-Sepharose (Amersham, 17-0780-01). Samples were incubated for 1.5 h at 4 °C on a rotator.

The resin was centrifuged at 1,000g for 2 min at 25 °C. The supernatants were removed and the resins were combined into two tubes and washed in series with the following buffers: twice with FA lysis buffer, twice with high-salt wash buffer, twice with FA wash 2 buffer, twice with FA wash 3 buffer, and once with TE (10 mM Tris-Cl, pH 8, 1 mM EDTA). For each wash, 15 ml of the indicated buffer was added, the sample was incubated for 5 min at 25 °C on a rotator, the resin was centrifuged at 1,000g for 2 min at 25 °C, and the supernatant was discarded. Additionally, the resin was transferred to a new tube during the second FA lysis and FA wash 3 washes to reduce background. Finally, the resin in each tube was resuspended in 5.85 ml ChIP elution buffer (25 mM Trizma, 2 mM EDTA, 0.2 M NaCl, 0.5% SDS) and incubated for 15 min at 65 °C. The resin was centrifuged and all eluates were combined in a single tube.

The crosslinks were reversed and the ChIP DNA was isolated and gel-purified as described previously<sup>1</sup> (see Supplementary Fig. 3), with the exception that during gel-purification DNA fragments ranging in size from 100 to 200 bp were excised. Gel-purified DNA was subject to pyrosequencing using the Roche GS20/FLX in accordance with the manufacturer's instructions. Raw sequencing reads can be accessed through NCBI Trace Archives TI SRA000283 under the Sequencing Center designation 'CCGB'. Bulk downloads or specific queries of nucleosome positions can be accessed from <http://atlas.bx.psu.edu/> or Supplementary Table 1.

**H2A.Z nucleosome mapping.** Sequencing reads were aligned to the FlyBase *D. melanogaster* reference genome (release 5.2) using BLAST. Sequences were aligned to all regions with >90% identity. Aligned regions were denoted on the *Drosophila* genome browser (<http://atlas.bx.psu.edu/>) by the coordinate of each read midpoint. Because the entire nucleosomal DNA was sequenced, both nucleosome borders were identified, thereby allowing nucleosome positions to be defined relative to each border: 73 bp interior to each end. The clustered distribution of reads (bar graph) were smoothed at a coarse-grain level, as described previously<sup>1</sup>, using a value of 20 as the 'smoothing parameter' (see Supplementary Fig. 2). The smoothing included adjustments of bar heights (read locations) in proportion to empirical determinations of MNase bias at each cut site (Supplementary Fig. 4). Nucleosome positions were defined as the closest genomic coordinate to the peak of the smoothed distribution. Three coarse-grain determinations were made using: only the Watson (W) strand, only the Crick (C) strand, and both strands. The error between the W and C calls is presented in Supplementary Fig. 7.

Analysis was performed on only those 112,750 nucleosomes that were defined by three or more reads, although virtually identical patterns (and conclusions) were achieved when all H2A.Z nucleosomes (defined by one or more reads) were included (data not shown). Nucleosome fuzziness (Supplementary Fig. 16) was calculated as the standard deviation for the set of reads for which midpoint locations (defined above) resided within 73 bp of the coarse-grain nucleosomal midpoint.

**Nucleosome distribution profiles.** The annotation of all *Drosophila* features (TSS included) were downloaded from FlyBase release 5.2 ([ftp://ftp.flybase.net/genomes/dmel/dmel\\_r5.2\\_FB2007\\_01/fasta](ftp://ftp.flybase.net/genomes/dmel/dmel_r5.2_FB2007_01/fasta)). The TSSs are annotated in the feature transcripts, and represent both experimental and computationally derived determinations. Only 3,419 (~24% of total) messenger RNA genes have alternative transcripts, many of which have the same TSS but with variant exon and intron structure. To remove the redundancy, only the 5'-most TSS was used in this study for genes with alternative transcripts. This eliminates TSSs located internal to genes. However, indistinguishable results were obtained when we used the 3'-most TSS for each gene (data not shown). The facts that *Saccharomyces*<sup>49,50</sup> and *Drosophila* TSS annotations were each derived from multiple laboratory sources, include both experimental and computational predictions, are internally consistent (yielding robust nucleosome landscape patterns), and are widely used, would indicate that there is unlikely to be a level of random or systematic error in the annotations that would lead to the offset of the nucleosome landscape reported in Fig. 1b.

In Fig. 1b, only those genes with a +1 H2A.Z nucleosome were examined, although essentially the same result was obtained when all genes were examined. In Fig. 4a, only those genes with an H2A.Z nucleosome within 1 kb of the ORF end were examined, although essentially the same result was obtained when all genes were examined.

Regions located internal to or within 300 bp of a nearby gene were removed from the analysis, except where indicated, to minimize potential influence from nearby genes. A minimum of 300 bp in either direction of the reference feature was analysed. None of these filters substantially affected the distribution of the data or conclusions. Equivalent filters were applied to the other data sets (described below). Nucleosome distances from the reference feature were binned in 10-bp intervals. Bin data were normalized to number of regions represented in each bin, and smoothed using a moving average. The size of the moving average (typically between three and eight bins) was set to provide the optimal balance between signal and noise for purposes of visual display. The size of moving average within the chosen range had no effect on the conclusions drawn.

For Fig. 5b, nucleosome positions for 1,956 'paused' genes were binned (10-bp bins) and plotted as a three-bin moving average. The number of genes used in the 'not-paused' class was 8,736. This number includes genes that either lack Pol II or lack a 'paused' Pol II. A paired *t*-test confirmed that the 'paused' pattern is not a subset of the 'not-paused' pattern (*P*-value,  $10^{-43}$ ), and that it is shifted from the 'not paused' pattern (*P*-value,  $10^{-9}$ ).

**Preparation of bulk nucleosomes.** Nucleosomal DNA from 3 g of Oregon R embryos (0–12 h old) was prepared similarly to H2A.Z nucleosomal DNA as described above, with the following change. After sonication, solubilized DNA was treated with an equal volume of 6 M urea, and was then dialyzed for 20–24 h against 100-fold volume of FA lysis buffer + 0.05% SDS at 4 °C (SpectraPor membrane, 6,000–8,000 molecular weight cut-off, 632650). This treatment enhanced the recovery of Pol II during immunoprecipitation (see below). A sample of this material was gel-purified (the remainder was subjected to Pol II immunoprecipitation), and was then amplified by ligation-mediated polymerase chain reaction (LM-PCR). The bulk nucleosomal DNA was amplified by LM-PCR and hybridized to Affymetrix GeneChip *Drosophila* tiling 1.0R arrays, each of which contains about 3 million oligonucleotide 25-bp probes that cover the euchromatic portion of the genome at an average resolution of about



one per 36 bp. Bulk nucleosome mapping data were analysed using the 'model-based analysis of tiling arrays' (MAT) program<sup>51</sup>, using the following settings: bandwidth 70 bp, minimum probe count 3, MaxGap 147 bp (which reflects the nucleosomal DNA length), 25-mer probe and 36 bp average probe separation. Interval cutoff was set to a  $P$ -value of  $1 \times 10^{-2}$ . We developed a peak detection algorithm that iteratively selects the maximal, non-overlapping peaks based on the MAT score assigned to the probe. The script is available on request. A total genome-wide count of 415,119 nucleosomes was obtained by these criteria (Supplementary Table 5).

**H2A.Z correlation to messenger RNA expression.** Data in Fig. 1c were previously normalized to 0–4 h expression to control for the presence of maternal RNA<sup>15</sup>. Shown is a moving median of 300 gene windows from 7,657 genes. Similar profiles were obtained when only the H2A.Z read count in the +1 nucleosome was considered, or if the genic read count was normalized to gene length (not shown).

**DNA motif distribution.** The 194 motifs were obtained by combining the novel motifs and known transcription factor (TF) motifs from a previous report<sup>23</sup>. Three motifs were excluded owing to rare occurrences (twin of eyeless recognition motif, TGGAGGDDGGWAHTMATBVRTGWDD DRKKMW; glass recognition motif, CAATGCACTTCTGGGGCTTCCAC; and abnormal chemosensory jump 6 recognition motif, TGCATAATTA ATTAC). Four short or highly degenerate motifs (prospero recognition motif, CWYBCY; apterous recognition motif, TAAT; mitochondrial transcription factor A recognition motif, TTATS; and bric a brac 1 recognition motif, WHWWWWWW-WWKK) caused computational failure owing to over-abundance. Their locations were therefore restricted to within  $\pm 1$  kb of the TSS.

On the basis of the motif consensus sequences, all occurrences of each motif were identified on both strands across the entire fly genome. Palindromic motifs were only searched on one strand. Motif midpoint distances to the TSS were calculated and binned (size = 10 bp). Bin counts were smoothed using a seven-bin moving average and plotted in Fig. 2b and Supplementary Fig. 12. Each motif was classified into one of four classes on the basis of visual inspection of the profiles (Supplementary Table 4), as shown in Fig. 2. To derive the values in Fig. 2b, bin counts were normalized to the average bin count from  $\pm 1$  kb of the TSS, and then  $\log_2$  transformed. The frequency of each motif distance from an H2A.Z nucleosome midpoint located throughout the entire genome was measured in the same manner. Several negative controls were performed as described in Supplementary Fig. 13.

**GAGAG site identification.** The *D. melanogaster* genome (r5.2) was scanned for GAGAG motifs that resided within 20 bp of another GAGAG motif (Supplementary Table 3), because they tend to co-occur<sup>43,52</sup>. Such clusters were binned in 20-bp intervals based on the distance from the TSS. Binned data were smoothed using a five-bin moving average. Clusters that were within 300 bp of an adjacent gene were removed from the analysis (which had little effect on the distribution).

**GAGA-associated factor binding.** ChIP–chip data for genome-wide GAF were obtained from ref. 30, and represent GAGA-associated factor (GAF) binding in Schneider 2 cells, as detected by Affymetrix *Drosophila* tiling 1.0R arrays. Signal analysis, interval analysis and peak calling were performed using MAT software<sup>51</sup>. The bandwidth value was 70 and the MaxGap value was 150 bp, and peaks were called using a 1% false-discovery rate threshold. The distance from the TSS was calculated for the 2,903 GAF peaks above a 1% false-discovery rate threshold. The GAF composite distribution profile was generated by binning these distance values in 20-bp intervals for all genes that contain a GAF peak within 1 kb of the transcription start site. The data were smoothed using an eight-bin sliding window average in one-bin steps.

**Dinucleotide frequencies.** To survey for all 16 dinucleotides across nucleosomal DNA, we used all 7,600 nucleosomal DNA sequences that were exactly 147 bp long. Dinucleotide counts were performed in the 5'–3' direction on both strands and were summed. Data were smoothed using a 3 bp moving average, except at position  $\pm 73$  and  $\pm 72$  at which 1 bp and 2 bp moving averages, respectively, were used.

**Nucleosome-positioning sequences.** Correlation profiles were calculated as described elsewhere<sup>25</sup>, using the AA/TT dinucleotide pattern obtained previously<sup>24</sup>. We also used a *Drosophila*-specific pattern in which the AA/TT dinucleotide pattern was derived from the top 1,000 most well positioned H2A.Z nucleosomes, representing <1% of all nucleosomes, and obtained essentially the

same result (data not shown). The CC/GG pattern was derived from the same nucleosome set used to derive the AA/TT pattern. This pattern is shown in Supplementary Fig. 15. The *Drosophila* pattern represents dinucleotide positional frequency profiles dyad-symmetrized as described previously<sup>24</sup>. In all cases, the number of nucleosomes used to define the search pattern represents <1% of all nucleosome locations, and therefore the nucleosomes used in the search pattern are expected to contribute <1% to resulting genome-wide correlation profiles.

**Genome-wide mapping of RNA polymerase II.** The genome-wide distribution of Pol II was determined as described elsewhere<sup>33</sup>. In brief, stage-14 embryos were crosslinked with formaldehyde; chromatin was isolated by CsCl<sub>2</sub> gradient ultracentrifugation and sonicated; and Pol II was immunoprecipitated using the 8WG16 antibody (Covance) against the C-terminal domain of the Pol II large subunit. The Pol II ChIP and control ChIP DNA samples, along with DNA purified from input chromatin samples, were amplified and hybridized to Affymetrix GeneChip *Drosophila* tiling 1.0R arrays. The ChIP–chip data were analysed using TiMAT—software developed by the Berkeley *Drosophila* Transcription Network Project<sup>33</sup>. The number of Pol II-crosslinked regions was 4,286, and were identified by genome-wide analysis at a 1% false discovery rate based on the symmetric null distribution method of TiMAT (Supplementary Table 6). These regions were ranked on the basis of the peak window score of each region. Within these regions, peaks of Pol II were located using the Mpeak algorithm<sup>54</sup>. All parameters for Mpeak were set at the default, except for the 'Largest search range', which was reduced to 1.0 kb. 1,956 genes were found to contain at least one Pol II Mpeak within 1 kb of a TSS, and no Pol II Mpeak between +1 kb and the end of the gene. These genes are defined as putative Pol II-paused genes. Pol II locations in Fig. 5a were binned in 20-bp intervals. Eight-bin moving averages were generated and plotted as a filled line graph. Essentially the same result was obtained when all genes were examined (not shown). 8,736 genes lacked a Pol II Mpeak anywhere in the transcription unit, of which 3,742 contained Pol II crosslinked regions somewhere in the body of the gene, and 4,994 lacked significant Pol II binding.

**Genome-wide mapping of nucleosomes bound to RNA polymerase II.** Material that was accumulated before gel purification of bulk nucleosomes (see above) was incubated for 16–18 h at 4 °C with 54  $\mu$ l anti-Rpb3 antibodies (raised in rabbit), and immunoprecipitated as described for H2A.Z. Immunoprecipitated DNA was not detected in the absence of antibody or in the absence of crosslinking (not shown). Eluate DNA was gel-purified in the size range of 75–200 bp, and 30 ng DNA was amplified by LM-PCR (primer sequence, GCGGTGACCCGGGAGATCTGAATTC). Two biological replicates were prepared for hybridization with the GeneChip WT Double-Stranded DNA Terminal Labelling Kit (Affymetrix 900812) using the manufacturer's recommended protocol, and hybridized to Affymetrix GeneChip *Drosophila* tiling 1.0R arrays. Nucleosome–Pol II interaction data were analysed using MAT, as described above for the bulk nucleosome analysis. A total genome-wide count of 82,969 Pol II-crosslinked nucleosomes were obtained using this criteria (Supplementary Table 7). The distribution of these nucleosomes relative to the TSS is displayed in Fig. 5c, and represents the average of two biological replicates. The nucleosome count data were binned in 10 bp and further smoothed using a three-bin moving average.

47. Fyrberg, E. & Goldstein, L. *Drosophila melanogaster*: Practical Uses in Cell and Molecular Biology. *Methods Cell Biol.* **44**, 1–732 (1994).
48. Orlando, V., Jane, E. P., Chinwalla, V., Harte, P. J. & Paro, R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* **17**, 5141–5150 (1998).
49. David, L. et al. A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA* **103**, 5320–5325 (2006).
50. Zhang, Z. & Dietrich, F. S. Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Res.* **33**, 2838–2851 (2005).
51. Johnson, W. E. et al. Model-based analysis of tiling-arrays for ChIP–chip. *Proc. Natl Acad. Sci. USA* **103**, 12457–12462 (2006).
52. van Steensel, B., Delrow, J. & Bussemaker, H. J. Genomewide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding. *Proc. Natl Acad. Sci. USA* **100**, 2580–2585 (2003).
53. Li, X. Y. et al. Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6**, e27 (2008).
54. Kim, T. H. et al. A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).



# Attributing physical and biological impacts to anthropogenic climate change

Cynthia Rosenzweig<sup>1</sup>, David Karoly<sup>2</sup>, Marta Vicarelli<sup>1</sup>, Peter Neofotis<sup>1</sup>, Qigang Wu<sup>3</sup>, Gino Casassa<sup>4</sup>, Annette Menzel<sup>5</sup>, Terry L. Root<sup>6</sup>, Nicole Estrella<sup>5</sup>, Bernard Seguin<sup>7</sup>, Piotr Tryjanowski<sup>8</sup>, Chunzhen Liu<sup>9</sup>, Samuel Rawlins<sup>10</sup> & Anton Imeson<sup>11</sup>

Significant changes in physical and biological systems are occurring on all continents and in most oceans, with a concentration of available data in Europe and North America. Most of these changes are in the direction expected with warming temperature. Here we show that these changes in natural systems since at least 1970 are occurring in regions of observed temperature increases, and that these temperature increases at continental scales cannot be explained by natural climate variations alone. Given the conclusions from the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report that most of the observed increase in global average temperatures since the mid-twentieth century is very likely to be due to the observed increase in anthropogenic greenhouse gas concentrations, and furthermore that it is likely that there has been significant anthropogenic warming over the past 50 years averaged over each continent except Antarctica, we conclude that anthropogenic climate change is having a significant impact on physical and biological systems globally and in some continents.

The IPCC Working Group II Fourth Assessment Report found, with very high confidence, that observational evidence from all continents and most oceans shows that many natural systems are being affected by regional climate changes, particularly temperature increases<sup>1,2</sup>. The Working Group II further concluded that a global assessment of data since 1970 shows that anthropogenic warming is likely (66–90% probability of occurrence) to have had a discernible influence on many physical and biological systems. Here we expand this assessment with a larger database of observed changes and extend the attribution from the global to the continental scale using multiple statistical tests. We also consider the part that other driving forces, especially land-use change, might have played at the study locations.

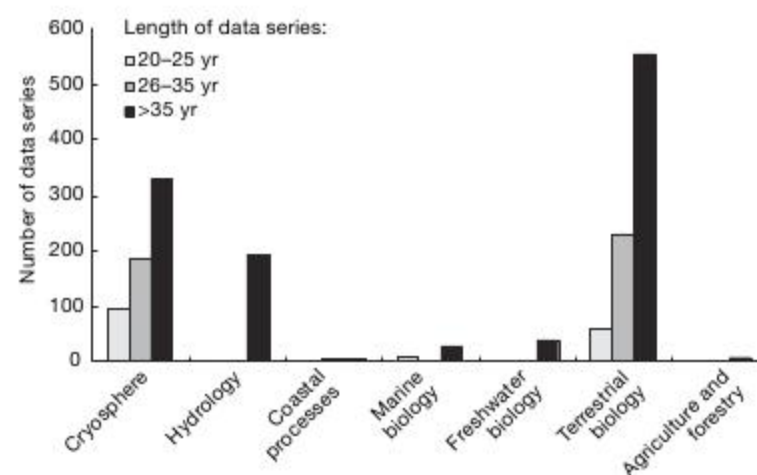
Observed responses to climate change are found across a wide range of systems as well as regions. Changes related to regional warming have been documented primarily in terrestrial biological systems, the cryosphere and hydrologic systems; significant changes related to warming have also been studied in coastal processes, marine and freshwater biological systems, and agriculture and forestry (Fig. 1). In each category, many of the data series are over 35 years in length.

Responses in physical systems include shrinking glaciers in every continent<sup>3,4</sup>, melting permafrost<sup>5,6</sup>, shifts in the spring peak of river discharge associated with earlier snowmelt<sup>7,8</sup>, lake and river warming with effects on thermal stratification, chemistry and freshwater organisms<sup>9–11</sup>, and increases in coastal erosion<sup>12–14</sup>. In biological systems, changes include shifts in spring events (for example, leaf unfolding, blooming date, migration and time of reproduction), species distributions and community structure<sup>15–18</sup>. Additionally, studies have demonstrated changes in marine-ecosystem functioning

and productivity, including shifts from cold-adapted to warm-adapted communities, phenological changes and alterations in species interactions<sup>19–22</sup>.

## Detection and attribution in natural systems

Following the definition of attribution of observed changes in the climate system<sup>23</sup>, changes in physical and biological systems are attributed to regional climate change based on documented



**Figure 1 | Data series of observed changes in physical and biological systems.** Length of the data series and types of observed changes in physical and biological systems. COST725 data series of terrestrial biological changes (>28,000 European phenological time series<sup>17</sup>) were measured over 30 years (1971–2000; not displayed).

<sup>1</sup>NASA/Goddard Institute for Space Studies and Columbia Center for Climate Systems Research, 2800 Broadway, New York, New York 10025, USA. <sup>2</sup>School of Earth Sciences, University of Melbourne, Victoria 3010, Australia. <sup>3</sup>School of Meteorology, University of Oklahoma, 100 East Boyd Street, Norman, Oklahoma 73019, USA. <sup>4</sup>Centro de Estudios Científicos, Avenida Arturo Prat 514, Casilla 1469, Valdivia, Chile. <sup>5</sup>Center of Life and Food Sciences Weihenstephan, Technical University of Munich, Am Hochanger 13, 85 354 Freising, Germany. <sup>6</sup>Stanford University, Center for Environmental Science and Policy, Stanford, California 94305, USA. <sup>7</sup>INRA Unité Agroclim, Site Agroparc, domaine Saint-Paul, F-84914 Avignon Cedex 9, France. <sup>8</sup>Department of Behavioural Ecology, Institute of Environmental Biology, Adam Mickiewicz University, Umultowska 89, PL-61-614 Poznań, Poland. <sup>9</sup>China Water Information Center, Lane 2 Baiguang Road, Beijing 100761, China. <sup>10</sup>Caribbean Epidemiology Center, 16–18 Jamaica Boulevard, Federation Park, PO Box 164, Port of Spain, Trinidad and Tobago. <sup>11</sup>3D-Environmental Change, Curtiuslaan 14, 1851 AM, Heiloo, Netherlands.



statistical analyses confirmed by process-level understanding in the interpretation of results. For example, a statistical association between poleward expansion of species' ranges and warming temperatures is expected when temperatures exceed physiological thresholds. The observed changes in both climate and the natural system are demonstrated to be: unlikely to be entirely due to natural variability; consistent with the estimated responses of either physical or biological systems to a given regional climate change; and not consistent with alternative, plausible explanations of the observed change that exclude regional climate change.

Attribution of changes in natural systems to anthropogenic warming requires further analysis because the observed regional climate changes must be attributed to anthropogenic causes. Combining these two types of attribution, called 'joint' attribution<sup>2</sup>, has lower statistical confidence than either of the individual attribution steps alone.

One approach to joint attribution, which uses what may be called an 'end-to-end' method, has already been conducted in several studies of specific physical and biological systems. This approach involves linking climate models with process-based or statistical models to simulate changes in natural systems caused by different climate forcing factors, and comparing these directly with observed changes in natural systems. When temperature data from the HadCM3 global climate model were used to examine the likely cause for changes in the timing of spring events of Northern Hemisphere wild animals and plants, results show the strongest agreement when the modelled temperatures were derived from simulations incorporating anthropogenic forcings<sup>24</sup>. Other similar studies have shown that the retreat of two glaciers in Switzerland and Norway cannot be explained by natural variability of climate and glacier mass balance<sup>25</sup>, that observed global and Arctic patterns of changes in streamflow are consistent with the response to anthropogenic climate change<sup>26,27</sup>, and that the observed increase in the area of forests burned in Canada over the last four decades is consistent with the response caused by anthropogenic climate change<sup>28</sup>.

Here we conduct a joint attribution study across multiple physical and biological systems at both the global and the continental scale. We demonstrate statistical consistency of observed changes (which are very unlikely to be caused by natural internal variability of the systems themselves or other driving forces) in natural systems with warming and conduct spatial analyses that show that the agreement between the patterns of observed significant changes in natural systems and temperature changes is very unlikely to be caused by the natural variability of the climate (Supplementary Fig. 1). Combined with the attribution of global and continental-scale warming to anthropogenic climate forcing demonstrated by IPCC Working Group I Fourth Assessment Report, this analysis provides strong support for joint attribution of observed impacts.

### Consistency with warming

Based on a database of documented responses in physical and biological systems from 1970 to 2004, temperature-related changes have been observed in all continents. Each documented response is a 'statistically significant' signal that is beyond the natural internal variability of those systems. The largest numbers of entries in the database are for Europe and North America, followed by North Central Asia (Fig. 2). Sparse evidence of responses related to temperature changes exists in Latin America, Africa and Australia. Physical and biological systems in regions without data series may or may not be changing, but are not documented in peer-reviewed literature.

Most (about 90% of the >29,500 data series,  $P \ll 0.001$ ) changes in these systems at the global scale have been in the direction expected as a response to warming. Ninety-five per cent of the 829 documented physical changes have been in directions consistent with warming, such as glacier wastage and an earlier spring peak of river discharge. For biological systems, 90% of the ~28,800 documented changes in plants and animals are responding consistently to temperature changes (mostly by means of earlier blooming, leaf

unfolding and spring arrival). Warming in oceans, lakes and rivers is also affecting marine and freshwater biological systems (for example, changes in phenology, migration and community composition in algae, plankton and fish).

An evaluation of possible publication bias has been undertaken using comprehensive phenological network data in Europe<sup>29</sup>, in which a systematic analysis of all available records (for example, leafing and flowering) documented the percentages of data series that are not changing and of significant changes in both directions (for example, in spring, in 66% there is no significant change, in 31% the onset dates are significantly advanced, and in 3% the onset dates are significantly delayed)<sup>29</sup>. The percentage of data series with significant changes consistent with warming found in Europe (~90%) is close to that found in North America and Asia, providing an indication that the database may represent an unbiased sample of changes in both directions in those continents.

### Spatial analyses at global and continental scales

The IPCC Working Group I Fourth Assessment Report concluded that most of the observed increase in global average temperatures since the mid-twentieth century is very likely (> 90% probability of occurrence) to be due to the observed increase in anthropogenic greenhouse gas concentrations<sup>30</sup>. It is very likely that the observed warming patterns cannot be explained by changes in natural external forcing factors, such as changes in solar irradiance or volcanic aerosols; the latter is likely to have had a cooling influence during this period.

At the global scale, agreement between the pattern of observed changes in physical and biological systems and the pattern of observed temperature change holds for two different gridded temperature data sets and two different pattern-comparison methods, and is exceptionally unlikely ( $P \ll 0.01$ ) to be explained by natural internal climate variability or natural variability of the systems; the latter is determined in the individual studies (Fig. 3). The spatial coherence of temperature trends across the globe is taken into account in these pattern comparisons using more than 3,000 years of climate model simulation data. The prevalence of observed statistically significant changes in physical and biological systems in expected directions consistent with anthropogenic warming in every continent and in most oceans means that anthropogenic climate change is having a discernible effect on physical and biological systems at the global scale.

For the first time, IPCC Working Group I Fourth Assessment Report extended its attribution of temperature trends to the continental scale, concluding that it is likely that there has been significant anthropogenic warming over the past 50 years averaged over each continent except Antarctica<sup>31</sup>. Similarly, a discernible anthropogenic influence is found in changes in natural systems in some continents where there is sufficient spatial coverage of responses in natural systems, including Asia and North America, and marginally in Europe. In these continents, there is a much greater probability of finding coincident significant warming and observed responses in the expected direction. Despite the presence of strong climate variability related to the North Atlantic Oscillation in Europe as well as its relatively small size, which makes it harder to distinguish signal from noise<sup>31</sup>, the plethora of evidence allows a signal to be detected, primarily in biological systems. The statistical agreement between the locations and directions of observed significant changes in natural systems and observed significant warming across Asia and North America ( $P < 0.05$ ) and across Europe ( $P \sim 0.1$ ) is very unlikely to be due to natural variability alone (Fig. 3). Responses not consistent with warming observed in  $5^\circ \times 5^\circ$  grid cells with warming temperature may be due to those systems responding to seasonal rather than recorded annual changes or to local cooling not represented in average cell temperatures; biological variation across species may also have a role (for example, late flowering species tend to be less affected by warming than earlier flowering ones). For the other continents,



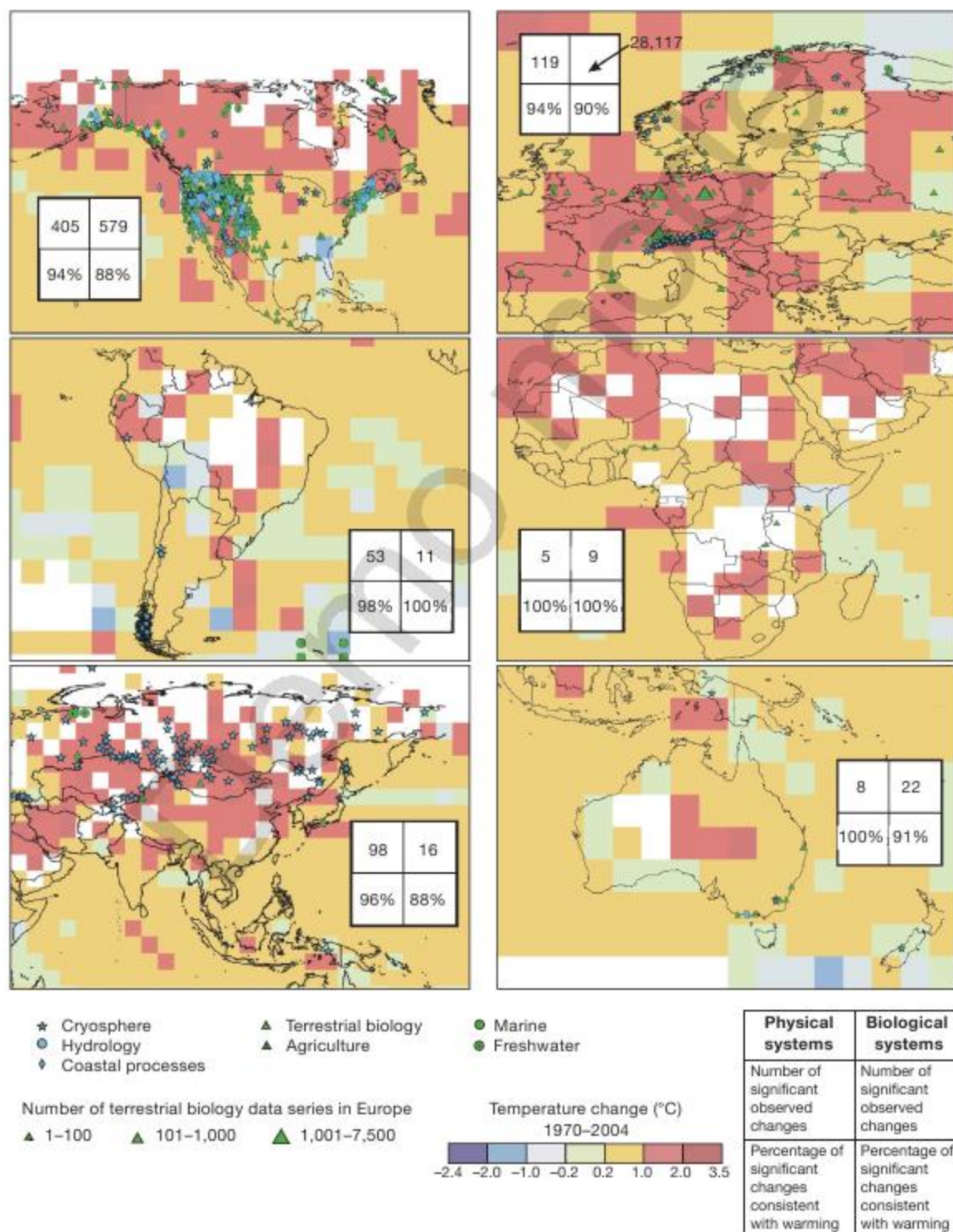
the sparse coverage of observed response studies makes it difficult to separate the observed responses related to anthropogenic temperature rise from those possibly caused by large-scale natural climate variations.

### Discussion and conclusions

The wide variety of observed responses to regional climate trends in expected directions combined with the attribution of climate trends to anthropogenic causes at both global and continental scales<sup>30</sup> demonstrates that anthropogenic climate change is already having a significant impact on multiple systems globally and in some continents. Most observed system changes are found in the cryosphere and in terrestrial biological systems and are consistent with the functional understanding and modelled predictions of climate change

impacts. The far fewer data series in Africa, Australia and Latin America are closely co-located with warming, but these cannot yet be attributed to anthropogenic climate forcing.

The issues of other climate and non-climate driving forces are important. In considering other drivers of change for phenology, much of the evidence in plants comes from changes observed in the spring. Even though day length can have a modulating effect on spring phenology depending on the plant species, it is not a factor in these studies because species remain *in situ* for the length of the time series, during which day length has not changed. There is also the possibility that increasing CO<sub>2</sub> is directly influencing plant phenology; however, experimental results show no consistent direction of response (that is, an advance or delay)<sup>32</sup>. Concerning trees, older trees tend to unfold leaves in spring later than younger



**Figure 2 | Location and consistency of observed changes with warming.** Locations of significant changes in physical systems (snow, ice and frozen ground as well as hydrology and coastal processes) and biological systems (terrestrial, marine and freshwater biological systems), and linear trends of

surface air temperature (HadCRUT3; ref. 35) between 1970 and 2004. Regions are based on data in refs 36 and 37. White areas do not contain sufficient climate data to estimate a trend. Note that there are overlapping symbols in some locations; Africa includes parts of the Middle East.



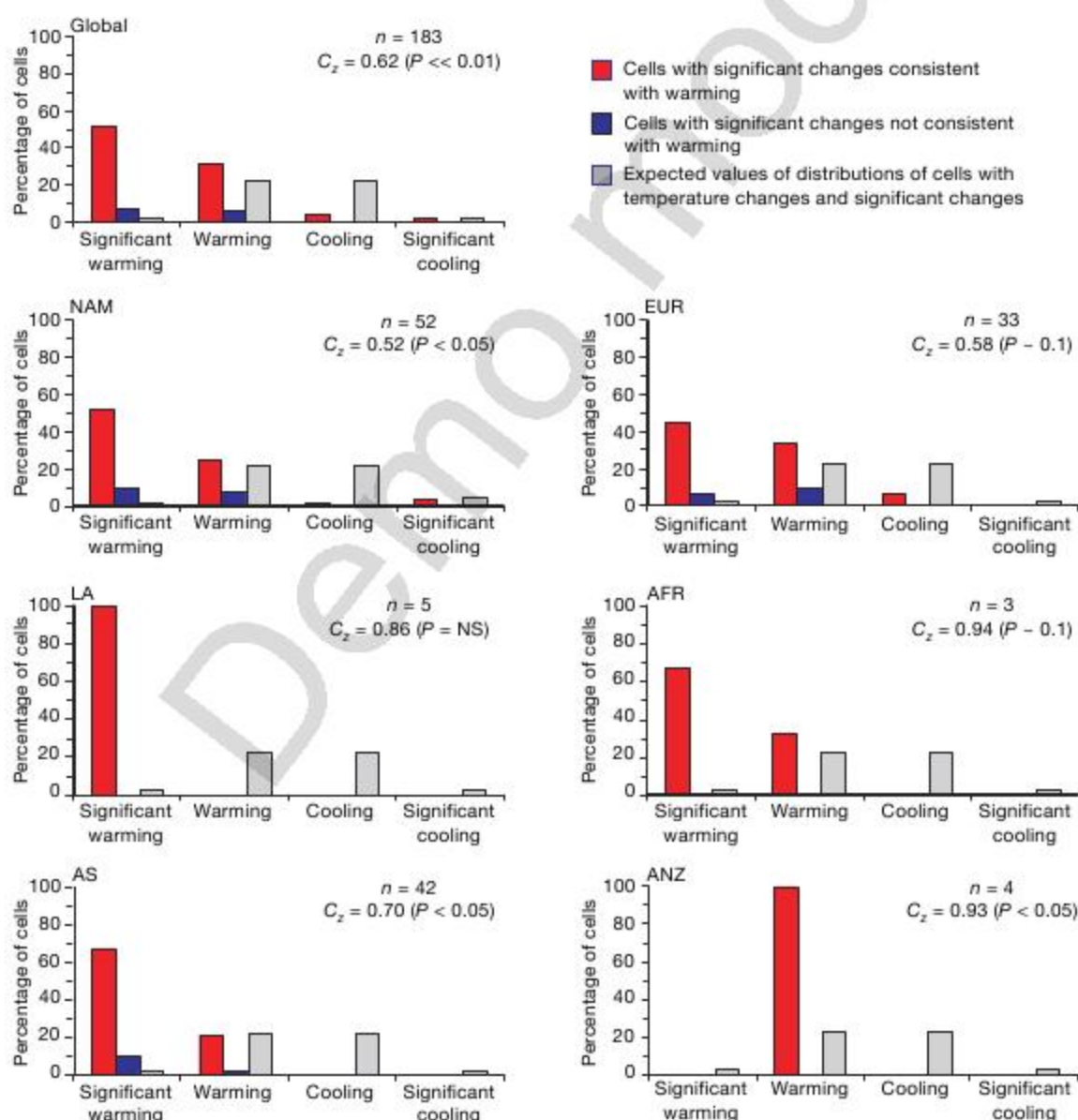
ones, so with longer time series on one specific object, the onset dates should become later with time owing to ageing, not earlier as observed owing to warming. Finally, some of the plant data, especially in Europe, come from phenological gardens that have been protected from the direct effects of land-use change for decades.

Land-use change, management practices, pollution and human demography shifts are all—along with climate—drivers of environmental change. Explicit consideration of these factors in observed-change studies strengthens the robustness of the conclusions. To determine the role of other driving forces in the data series used in this analysis, we assessed the likelihood of their having a direct effect on the observed system (see Supplementary Table 1). Out of the ~29,500 data series documented in ~80 studies included in the database, effects documented in only 3 studies (9 data series in 4 cells) were likely to have been caused by a driving force other than climate change (for example, habitat destruction, pollution or fishery by-catch disposal). Removing these data series from the statistical analyses does not change the results significantly.

Land-use change can affect physical and biological systems indirectly through its effects on climate. Yet, for recent climate trends on a global scale, the effect of land-use change is small<sup>31</sup>. In addition,

because these effects may result in warming in some regions and cooling in others (for example, agricultural expansion tends to warm the Amazon and cool the mid-latitudes)<sup>33,34</sup>, they are very unlikely to explain the coherent responses that have been found across the diverse range of systems and across the continental and global scales considered (Supplementary Table 2). Cooling in temperate regions occurs because the clearing of forests for agriculture may increase albedo during periods of snow cover, although recent afforestation may be dampening this effect.

Documentation of observed changes in physical and biological systems in tropical and subtropical regions is still sparse. These areas include Africa, South America, Australia, Southeast Asia, the Indian Ocean and some regions of the Pacific. One reason for this lack of documentation might be that some of these areas do not have pronounced temperature seasons, making events such as the advance of spring phenology less relevant. Other possible reasons for this imbalance are a lack of data and published studies, lag effects in responses, and resilience in systems. Improved observation networks are urgently needed to enhance data sets and to document sensitivity of physical and biological systems to warming in tropical and subtropical regions, where many developing countries are located.



**Figure 3 | Distribution of cells with temperature changes and significant observed changes.** Expected and observed distributions of cells with significant responses consistent with warming and distributions of cells with significant responses not consistent with warming for  $5^\circ \times 5^\circ$  grid cells of temperature change between 1970 and 2004 (HadCRUT3). The global total includes polar regions and marine systems. Shown is the number of cells (n)

with observed impacts and temperature data, the pattern congruence between locations of significant responses and standardized temperature trends ( $C_2$ ), and the probability ( $P$ ) that pattern agreement could be explained by natural internal variability of temperature fields. Abbreviations: AFR, Africa; ANZ, Australia and New Zealand; AS, Asia; EUR, Europe; LA, Latin America; NAM, North America; NS, not significant.



## METHODS SUMMARY

We developed a database of observed changes in natural systems from peer-reviewed papers, demonstrating a statistically significant trend in change in either direction related to temperature and containing data for at least 20 years between 1970 and 2004. Observations in the studies were characterized as a 'change consistent with warming' or a 'change not consistent with warming'. The databases of the observed significant changes in the natural systems were overlaid with two gridded observed temperature data sets and the spatial patterns of the observed system changes were compared with the observed temperature trends using two different pattern-comparison measures.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 January; accepted 19 March 2008.

1. IPCC in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) 7–22 (Cambridge Univ. Press, Cambridge, UK, 2007).
2. Rosenzweig, C. et al. in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) 79–131 (Cambridge Univ. Press, Cambridge, UK, 2007).
3. Dyurgerov, M. B. & Meier, M. F. in *Occasional Paper No. 58* (Institute of Arctic and Alpine Research, Univ. Colorado at Boulder, 2005).
4. Oerlemans, J. Extracting a climate signal from 169 glacier records. *Science* **308**, 675–677 (2005).
5. Frauenfeld, O. W., Zhang, T., Barry, R. G. & Gilichinsky, D. Interdecadal changes in seasonal freeze and thaw depths in Russia. *J. Geophys. Res.* **109**, D05101, doi:10.1029/2003JD004245 (2004).
6. Yoshikawa, K. & Hinzman, L. D. Shrinking thermokarst ponds and groundwater dynamics in discontinuous permafrost near Council, Alaska. *Permafrost Periglac. Process.* **14**, 151–160 (2003).
7. Cayán, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M. & Peterson, D. H. Changes in the onset of spring in the western United States. *Bull. Am. Meteorol. Soc.* **82**, 399–415 (2001).
8. Mote, P. W., Hamlet, A. F., Clark, M. P. & Lettenmaier, D. P. Declining mountain snowpack in western north America. *Bull. Am. Meteorol. Soc.* **86**, 39–49 (2005).
9. O'Reilly, C. M., Alin, S. R., Plisnier, P. D., Cohen, A. S. & McKee, B. A. Climate change decreases aquatic ecosystem productivity of Lake Tanganyika, Africa. *Nature* **424**, 766–768 (2003).
10. Sorvari, S., Korhola, A. & Thompson, R. Lake diatom response to recent Arctic warming in Finnish Lapland. *Glob. Change Biol.* **8**, 171–181 (2002).
11. Daufresne, M., Roger, M. C., Capra, H. & Lamouroux, N. Long-term changes within the invertebrate and fish communities of the Upper Rhone River: effects of climatic factors. *Glob. Change Biol.* **10**, 124–140 (2004).
12. Beaulieu, N. & Allard, M. The impact of climate change on an emerging coastline affected by discontinuous permafrost: Manitoounuk Strait, northern Quebec. *Can. J. Earth Sci.* **40**, 1393–1404 (2003).
13. Forbes, D. L., Parkes, G. S., Manson, G. K. & Ketch, L. A. Storms and shoreline retreat in the southern Gulf of St. Lawrence. *Mar. Geol.* **210**, 169–204 (2004).
14. Orviku, K., Jaagus, J., Kont, A., Ratas, U. & Rivas, R. Increasing activity of coastal processes associated with climate change in Estonia. *J. Coast. Res.* **19**, 364–375 (2003).
15. Root, T. L. et al. Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60 (2003).
16. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
17. Menzel, A. et al. European phenological response to climate change matches the warming pattern. *Glob. Change Biol.* **12**, 1969–1976 (2006).
18. Parmesan, C. Ecological and evolutionary responses to recent climate change. *Ann. Rev. Ecol. Evol. System.* **37**, 637–669 (2006).
19. Richardson, A. J. & Schoeman, D. S. Climate impact on plankton ecosystems in the Northeast Atlantic. *Science* **305**, 1609–1612 (2004).
20. Edwards, M. & Richardson, A. J. Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature* **430**, 881–884 (2004).
21. Beaugrand, G. & Reid, P. C. Long-term changes in phytoplankton, zooplankton and salmon related to climate. *Glob. Change Biol.* **9**, 801–817 (2003).
22. Atkinson, A., Siegel, V., Pakhomov, E. & Rothery, P. Long-term decline in krill stock and increase in salps within the Southern Ocean. *Nature* **432**, 100–103 (2004).
23. Mitchell, J. F. B. et al. in *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (ed. Houghton, J. T.) 695–738 (Cambridge Univ. Press, Cambridge, UK, 2001).
24. Root, T. L., MacMynowski, D. P., Mastrandrea, M. D. & Schneider, S. H. Human-modified temperatures induce species changes: joint attribution. *Proc. Natl Acad. Sci. USA* **102**, 7465–7469 (2005).
25. Reichert, B. K., Bengtsson, L. & Oerlemans, J. Recent glacier retreat exceeds internal variability. *J. Clim.* **15**, 3069–3081 (2002).
26. Milly, P. C. D., Dunne, K. A. & Vecchia, A. V. Global pattern of trends in streamflow and water availability in a changing climate. *Nature* **438**, 347–350 (2005).
27. Wu, P., Wood, R. & Stott, P. Human influence on increasing Arctic river discharges. *Geophys. Res. Lett.* **32**, L02703 (2005).
28. Gillett, N. P., Weaver, A. J., Zwiers, F. W. & Flannigan, M. D. Detecting the effect of climate change on Canadian forest fires. *Geophys. Res. Lett.* **31**, L18211, doi:10.1029/2004GL020876 (2004).
29. Menzel, A., Sparks, T., Estrella, N. & Roy, D. B. Geographic and temporal variability in phenology. *Glob. Ecol. Biogeogr.* **15**, 498–504 (2006).
30. IPCC in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. D. et al.) (Cambridge Univ. Press, Cambridge, UK, 2007).
31. Hegerl, G. C. et al. in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Solomon, S. et al.) 663–745 (Cambridge Univ. Press, Cambridge, UK, 2007).
32. Asshoff, R., Zott, G. & Korner, C. Growth and phenology of mature temperate forest trees in elevated CO<sub>2</sub>. *Glob. Change Biol.* **12**, 848–861 (2006).
33. Feddema, J. et al. A comparison of a GCM response to historical anthropogenic land cover change and model sensitivity to uncertainty in present-day land cover representations. *Clim. Dyn.* **25**, 581–609 (2005).
34. Bounoua, L., DeFries, R., Collatz, G. J., Sellers, P. & Khan, H. Effects of land cover conversion on surface climate. *Clim. Change* **52**, 29–64 (2002).
35. Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B. & Jones, P. D. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.* **111**, D12106, doi:10.1029/2005JD006548 (2006).
36. Giorgi, F. Variability and trends of sub-continental scale surface climate in the 20th century. Part I: observations. *Clim. Dyn.* **18**, 675–691 (2002).
37. Stott, P. A. Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.* **30**, 1728 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Palutikof, D. Rind and A. Watkinson for their feedback, and J. Mendoza for work on the graphics. The Goddard Institute for Space Studies authors acknowledge the support of the Earth Science Division, NASA Science Mission Directorate. D.K. is supported by the Australian Research Council as a Federation Fellow. Q.W. is supported by a Gary Comer Science and Education Foundation Postdoctoral Fellowship and by the National Science Foundation grant ATM-0555326. We acknowledge the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the multi-model data set. Support of this data set is provided by the Office of Science, US Department of Energy.

**Author Contributions** C.R., D.K., G.C., A.M., T.L.R., B.S., P.N. and M.V. conceived the analytical framework; P.N., M.V., A.M. and N.E. constructed the database; M.V., D.K. and Q.W. performed the statistical analyses; G.C., A.M., T.L.R., P.T., B.S., C.L. and S.R. provided expertise in observed changes in physical and biological systems; and P.N., A.M., C.R. and A.I. analysed other driving forces.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.R. ([crosenzweig@giss.nasa.gov](mailto:crosenzweig@giss.nasa.gov)).



## METHODS

**Database of observed changes.** We developed a database of observations from peer-reviewed papers (primarily published since the IPCC Third Assessment Report<sup>38</sup>), specifically documenting the data series in terms of system, region, longitude and latitude, dates and duration, statistical significance, type of impact, and whether or not land use was identified as a driving factor (see Supplementary Table 1). Data for the system changes were taken from ~80 studies (of which ~75 are new since the Third Assessment Report) containing >29,500 data series. Studies were selected that demonstrate a statistically significant trend in change in either direction in systems related to temperature or to other climate change variables as described by the authors, and that contain data for at least 20 years between 1970 and 2004 (although study periods may extend earlier or later). Observations in the studies were characterized as a 'change consistent with warming' or a 'change not consistent with warming'.

**Spatial analysis.** Databases of the observed significant changes in the natural systems and the regional temperature trends over the period 1970–2004 were overlaid in a geographical information system. For Europe, even though there were very large numbers of observed response data series in some cells, these were counted as single cells in the spatial analysis. Two different gridded observed temperature data sets were used: HadCRUT3 (ref. 35) and GHCN-ERSST (ref. 39), both of which were used in the IPCC Fourth Assessment Report. In each 5° × 5° grid cell, the observed system responses were assessed as consistent with warming or not consistent with warming—based on a decision rule of 80% or more of data series consistent with warming within a cell—providing a binary pattern of 183 (HadCRUT3) and 203 (GHCN-ERSST) cells across the globe. There are fewer cells with temperature data in the HadCRUT3 data set because it does not use any infilling of data from adjacent cells, unlike GHCN-ERSST. All cells with observed temperature data are included from each of the data sets, irrespective of the sign of the temperature trend.

The spatial patterns of the observed system changes were compared with the observed temperature trends using two different pattern-comparison measures. To assess the significance of these observed measures of pattern agreement, global temperature trend data were obtained from long control simulations with seven different climate models from the WCRP CMIP3 multi-model database at PCMDI, to represent the range of 35-year temperature trends across the globe resulting from natural climate variations. Details of the different models used are included in Supplementary Table 3. The global temperature trend fields from the climate models represent the spatial coherence and decadal variability of natural internal temperature variations.

Two different pattern-comparison measures were used: a binary pattern congruence (unc centred pattern correlation) between the gridded binary field of system responses consistent (or not consistent) with warming and the gridded field of positive (or negative) temperature trends; and a pattern congruence between the gridded binary field of system responses and the gridded field of standardized temperature trends (the 35-year temperature trends divided by the standard deviation of 35-year temperature trends caused by natural internal climate variations). For each of these measures, the observed values for the two different observed temperature-trend data sets were compared with the distributions obtained using temperature trends caused by natural internal climate variability, as represented by the climate models. Significant attribution was assigned when both spatial statistics methods and both temperature data sets showed significant results. Detailed results are presented in the Supplementary Information and are summarized in the section 'Spatial analyses at global and continental scales' above.

38. IPCC (ed.) *Climate Change 2001: Impacts, Adaptation, and Vulnerability: Contribution of Working Group II to the Third Assessment Report to the International Panel on Climate Change* (Cambridge Univ. Press, Cambridge, UK, 2001).
39. Smith, T. M. & Reynolds, R. W. A global merged land and sea surface temperature reconstruction based on historical observations (1880–1997). *J. Clim.* **18**, 2021–2036 (2005).



# naturejobs

## JOBS OF THE WEEK

**T**he much celebrated rise of interdisciplinary science has an intuitive appeal. Institutions have been foolish and shortsighted, goes the rationale. Disciplines have been isolated for too long. Discoveries happen at the margins, where biological principles need insights from physicists and chemistry problems require input from biologists. Flouting convention is key.

Yet some are sceptical about the true value of interdisciplinary centres. They argue that interdisciplinary is little more than a buzzword and that, as Rogers Hollingsworth, a science historian at the University of Wisconsin-Madison, put it: "Large research organizations have an enormous amount of inertia, and individuals have a great vested interest in the way they were trained, and what they were doing yesterday" (see *Nature* **451**, 872–873; 2008).

As part of an infrastructure 'arms race', institutions run the risk of building cutting-edge interdisciplinary facilities to attract attention, funds and talent, but that produce little sound or novel science. Peer-review panels may not so readily recognize interdisciplinary accomplishments; tenure committees may turn away candidates who have not specialized in a traditional field. As this week's feature details, young scientists looking for interdisciplinary training face various challenges, despite new graduate-school programmes (see page 422).

But this view presumes that the primary track for young interdisciplinarians is academia. This is increasingly not the case. The number of academic positions has fallen in many countries, whether interdisciplinary or otherwise, yet unemployment for scientists is at an all-time low. Those with science degrees are readily employable outside academia (see *Nature* **452**, 777; 2008). Positions in policy, industry and elsewhere need scientists; and a knowledge of how fields intersect can be advantageous. Increasingly, societal problems that are linked to science — such as alternative fuels or vaccine development — require interdisciplinary points of view. Alumni of interdisciplinary centres may have ample opportunities. If so, such centres could prove to be much more than a fad.

**Gene Russo is editor of Naturejobs.**

### CONTACTS

**Editor:** Gene Russo

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Andy Douglas (4975)  
e-mail: [a.douglas@nature.com](mailto:a.douglas@nature.com)  
**Business Development Manager:**  
Amelie Pequignot (4974)  
e-mail: [a.pequignot@nature.com](mailto:a.pequignot@nature.com)  
**Natureevents:**  
Claudia Paulsen Young (+44 (0) 20 7014 4015)  
e-mail: [c.paulsenyoung@nature.com](mailto:c.paulsenyoung@nature.com)  
**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)  
**Southwest UK/RoW:** Nils Moeller (4953)

**Scandinavia/Spain/Portugal/Italy:**  
Evelina Rubio-Hakansson (4973)  
**Northeast UK/Ireland:**  
Matthew Ward (+44 (0) 20 7014 4059)  
**North Germany/The Netherlands:**  
Reya Silao (4970)  
**South Germany/Austria:**  
Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**  
Stephen Russell  
To send materials use London address above.  
Tel: +44 (0) 20 7843 4816  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)  
**Naturejobs web development:** Tom Hancock  
**Naturejobs online production:** Dennis Chu

**US Head Office, New York**  
75 Varick Street, 9th Floor,  
New York, NY 10013-1917  
Tel: +1 800 989 7718

Fax: +1 800 989 7103  
e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**India**  
Vikas Chawla (+91 1242881057)  
e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**  
Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843  
Tel: +81 3 3267 8751  
Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**  
Ayako Watanabe (+81 3 3267 8765)  
e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)  
**Business Development Manager, Greater China/Singapore:**  
Gloria To (+852 2811 7191)  
e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)



# ASSEMBLY WORK

Graduate students can be key links in interdisciplinary science, but training them for this role is a challenge, says **Brian Vastag**.

**A**s a master's student in parasitology, Adriana Troyo became intrigued by dengue fever, a mosquito-borne disease that periodically flares up in her native Costa Rica. Troyo wanted to understand how geographical variation — such as urban tree cover — affected the disease's spread. She needed a PhD, but she wasn't sure which academic department would be best as her interests lie at the intersection of microbiology, epidemiology and geography.

So she entered an interdepartmental PhD programme at the University of Miami in Florida. There, she helped map ground-level mosquito data onto high-resolution satellite maps. The resulting 'risk maps' help guide country-wide mosquito-control efforts, says Troyo, now a professor of parasitology at the University of Costa Rica in San Jose.

For Troyo, becoming an interdisciplinary scientist was a natural move. But for many young scientists, boundary-busting can be daunting. Graduate schools and centres are still looking for the right ways to train a new generation of interdisciplinary scientists who can both speak the language of multiple fields and maintain enough expertise to take on cutting-edge problems.

## Beyond the buzzword

'Interdisciplinary' has been a buzzword in science for at least 20 years. All manner of collaborations have sprung up — from pairs of researchers joining forces to state-of-the-art dedicated facilities, such as Stanford University's Bio-X centre in Palo Alto, California, Arizona State University's Biodesign Institute in Tempe and Janelia Farm, a Howard Hughes Medical Institute centre that opened in Maryland in 2006.

That same year, the University of Manchester, UK, opened its Interdisciplinary Biocentre (MiB), a US\$68-million building able to house 75 research groups and more than 600 staff. University College London's CoMPLEX programme, launched in 1998, is moving into a new building in May. In Germany, the Max Planck institutes' interdisciplinary efforts include an Institute for Dynamics of Complex Technical Systems in Magdeburg.

According to an unpublished survey, 25–30% of science and engineering faculty members at major US research universities report an affiliation with an interdisciplinary venture of some sort. Survey author Barry Bozeman, a professor of public policy at the University of Georgia in Athens, is trying to understand how working at interdisciplinary centres affects researchers' career trajectories. This is important because, until recently, administrators and researchers

had learned little about how these centres operate, says Diana Rhoten, a programme director in the Office of Cyberinfrastructure of the US National Science Foundation (NSF).

Rhoten scrutinized the operations of six NSF-funded interdisciplinary centres. Her 2003 report, *A Multi-method Analysis of the Social and Technical Conditions for Interdisciplinary Collaboration*, concluded, in part, that graduate students were the 'bridges' between disciplines. This bridging was particularly stark at one centre: remove the graduate students from the network map and only single-discipline islands remained. "Pre-existing networks of scientists tend to cluster," says Rhoten. She is now wants to find out whether graduate students are merely running errands between faculty members, or whether they are really asking transdisciplinary questions and sparking collaborations. "I tend to believe it's a combination of both," says Rhoten. Young scientists are probably driving a lot of interdisciplinary collaborations.

Mike Wininger is a good example. He studied maths and physics as an undergraduate. But as a graduate



M. REYNOLDS/EPA/CORBIS



student he says he “wanted something more than just solving equations for the sake of solving equations — a more human element”. So Wininger, a PhD student at Rutgers University, New Jersey, moved into biomedicine. He is now designing an image-processing system to sort healthy embryonic stem cells from unhealthy cells, bringing mathematical rigour to the process.

Two years of Wininger's graduate education were funded by the NSF's flagship interdisciplinary training programme, the Integrative Graduate Education and Research Traineeship (IGERT). The programme was launched in 1997 to “catalyse a change in graduate education”. Some 150 colleges and universities hold five-year IGERT grants to train the next generation of interdisciplinary scientists. The programme is extremely competitive, with only about 5% of applying institutions winning funds, says programme director Carole Van Hartesveldt. A 2006 evaluation concluded that IGERT graduates had broader skill sets and were more oriented towards identifying and solving specific problems — advantages when hunting for jobs in industry.

### Taking a risk

Arizona State University is striving to create a model even more innovative and potentially risky than IGERT, according to Neal Woodbury, co-director of the new interdisciplinary graduate programme there. Students in the programme, which welcomes its first class this autumn, will have no home department. And, modelled after an approach taken by the university's Biodesign Institute, studies are organized around pragmatic problems — for example, how to devise better renewable fuels — rather than a canon of science facts. Students will also be involved in seeking funding to develop their work. Woodbury says that, based on pilot interdisciplinary programmes at the university, this ‘biological design’ PhD is likely to attract applicants who would never have pursued a traditional PhD — for example, individuals sick of working in industry but who still want to solve big problems. The dozen students in the first class have backgrounds ranging from bioengineering to chemistry to physics.

Initial overview courses in interdisciplinary programmes are key. Students from different disciplines must quickly get up to speed in each other's fields. At the MiB, which has tracks in molecular bioengineering, systems biology and biological mechanism and catalysis, students in a four-year PhD programme — a year longer than the typical UK doctorate — initially have several months of lectures across fields such as mathematical modelling, analytical science and informatics. Second-year student Shichina Kannambath says that the first year of training was an essential base for embarking on her three-year PhD project, an imaging analysis of transcription factors in eukaryotes. The MiB has graduate programme funding only for the systems biology track, but its director, John McCarthy, hopes to get funding for the other tracks as well.



Diana Rhoten (top) and Barry Bozeman are trying to assess how interdisciplinary centres really work.

Richard Zare, chair of Stanford University's chemistry department and a long-time interdisciplinary, recalls his bitter experience in the 1960s as a young professor at the University of Colorado with joint appointments in physics and chemistry. “As an untenured faculty member, it was a trap,” he says. “Each department regarded me in some sense as a spy from the other.” He left after a few years. Although interdisciplinary work is gaining acceptance, old disciplinary silos still can impede and even discourage interdisciplinary aspirations.

### Obstacles ahead

The risks of crossing departmental boundaries can be greatest for fledgling scientists. Tenure review committees can be a big obstacle, says Bozeman, as promotion criteria often fail to register interdisciplinary achievements. If a young researcher joins an interdisciplinary research centre, the fruits of such labour — as measured by publications, for instance — may not be readily apparent. “There is a sort of double jeopardy that goes on,” says Alan Johnston, deputy director of the CoMPLEX programme. Physicists, for example, might not see the point of the biological aspects and biologists might not see the point of the physics elements, he says.

Arizona State's Woodbury, on the other hand, plays down the risks of a poor publication record for interdisciplinary students. Ideally, the students still specialize on some part of the interdisciplinary project, while staying part of the team, thus giving them a core expertise. “But you do have to train students not to take that problem and go into a corner and never come back,” Woodbury says. When studying the NSF centres, Rhoten found that researchers reported publications as a less important benefit than the intellectual change they were experiencing.

Woodbury is, however, a bit wary of how a PhD in biological design might be perceived by employers. “We might have to modify our approach on the basis of how the students do,” he says. And Johnston notes that the shifts in research discipline that many interdisciplinary students undergo can make it harder to find postdoc positions.

McCarthy is concerned about finding even-handed interdisciplinary peer review among funding bodies. He and others have commissioned a report from the Royal Society of Chemistry on the topic, focusing on the interface between chemistry and biology. Due to be released soon, the report finds in part that UK funding bodies have improved in this regard, but could get better. The US National Institutes of Health is considering forming review panels geared to evaluating interdisciplinary research proposals.

Regardless, graduate students are likely to continue to act as the cement in interdisciplinary partnerships. “They don't know enough initially to be afraid of all the disciplinary problems of making programmes like this work,” says Woodbury.

**Brian Vastag is a freelance science writer based in Washington DC.**



A new breed: the Manchester Interdisciplinary Biocentre.



# MOVERS

**Craig Hogan, director, Center for Particle Astrophysics, Fermi National Accelerator Laboratory and professor of astronomy and astrophysics, University of Chicago, Illinois**



**1993–2008** Professor, astronomy and physics departments, University of Washington, Seattle, Washington  
**2002–05** Vice-provost for research, University of Washington, Seattle  
**1995–2001** Chair, astronomy department, University of Washington, Seattle

Craig Hogan was hooked on astrophysics the minute he learned that remnant heat from the Big Bang was still detectable. His wide-ranging contributions to the field include the co-discovery of 'dark energy' — the mysterious force behind the acceleration of the expanding Universe. But he hopes that future experiments will reveal an as-yet undetected dimension of the Universe.

"Craig has forged unusually original and versatile theoretical insights into astrophysics," says Martin Rees, Hogan's PhD adviser at the University of Cambridge, UK. "If you look at any number of subjects — from dark energy to how the Universe began — you'll find the earliest papers are from Craig."

After postdocs at the universities of Cambridge and Chicago and at the California Institute of Technology in Pasadena, Hogan helped to build the first theoretical group at the University of Arizona's Steward Observatory. There, he learned the inner workings of telescope-based experiments as they applied to theory projects.

Hogan eventually joined the physics and astronomy department at the University of Washington in Seattle, where it was easier to bridge his interest in those two fields. Despite the clouds and mountains, Washington had a telescope large enough to survey supernovae, key to detecting dark energy. Hogan plays down the discovery. "The tension between the age of the Universe and the velocity of the galaxies had suggested a cosmological constant for a long time," he says. The real surprise, he adds, was how well the experiment actually worked.

Hogan's research focus now includes proposed space-based experiments such as the Laser Interferometer Space Antenna (LISA), designed to detect gravitational waves in space. "So far we've turned snapshots of the Universe into a silent movie," says Hogan. "Detecting gravitational waves would be like adding a soundtrack — and that is potentially much more transformative to science as a whole than discovering dark energy."

As director of Fermilab's Center for Particle Astrophysics, Hogan will push for LISA as well as for other new ways to explore the physics of gravity and space-time. Fermilab will need a new focus once the Large Hadron Collider in Switzerland makes the lab's Tevatron particle collider obsolete. "Fermilab is the premier high-energy physics lab in the United States," says Hogan, "and gravity is the one force of nature it hasn't yet studied."

Virginia Gewin

## NETWORKS & SUPPORT

### The postdoc interview

The following questions should help you get the most out of your postdoc interview and, with luck, out of your postdoc experience as well.

- If the project is independent, will you be able to take it with you when the fellowship ends? Be wary of a 'conditional project' — if you take on project X first, you will get project Y.
- Will you be granted time to participate in a postdoc council, take courses, learn new technologies, improve communication skills, and so on? Does the mentor/principal investigator (PI) have funds for attending scientific meetings?
- How is authorship handled? How often and where does the lab publish?
- Where is the mentor along the tenure-track timeline? Senior PIs with productive track records are safer. But junior faculty members may be more eager to get more publications out.
- Will the mentor help you apply for small grants or fellowships? How stable is the current funding?
- Is your potential mentor receptive to collaborations with other labs?
- How are research supplies acquired and financed?
- Will you be able to meet lab members to talk freely about the lab and the mentor? A PI's pledges in the courtship phase might change once a commitment is made.

- Can you contact lab alumni? If so, ask them for the five best and five worst things about working with the mentor.
- Is productivity more important than the number of hours you work?
- What is the vacation schedule? Is there sick time? Is there a maternity/paternity policy or leave?
- Does the mentor hold regular lab or individual meetings? An individual development plan helps answer these questions in a written, contractual format.

- Does the institution have a postdoc policy? If so, read it in detail. Are there postdoc term limits?
- Does the institution have a postdoc office or association? If so, does it review all the offer letters to ensure the terms and stipend levels are fair?

If you walk away thinking, "If I could change one or two major things, it would be great", then keep on walking. Reflect on the interview, talk to your mentors, and trust your gut. ■

**Kryste Ferguson is an academic coordinator at the University of Pennsylvania's Office of Biomedical Postdoctoral Programs. Ivonne Vidal Pizarro is a programme administrator at the American Association for Cancer Research. Reprinted by permission of the Society for Advancement of Chicanos and Native Americans in Science.**

#### POSTDOC JOURNAL

### Judging me, judging you

"Good morning," I begin. "It's great to be here presenting my latest research." And it is good to be in front of faculty members and students, despite my sickening nerves. This is judgement day for getting a permanent faculty position.

Someone rustles my handout as I unveil my work, 'Invasions in heterogeneous environments'. Maybe handouts weren't such a good idea after all. Can I convey my interest, inspire people? Much as in a student lecture, I scan faces for some feedback. Nothing. So I catapult everyone into my favourite results.

The interview gets altogether more personal. How would I teach ecological theory? With whom would I collaborate? What resources would I require? Now, I've done my homework. I've spent evenings reading the department's website. I know their research, their courses, their 'future strategic plan', the funding opportunities, and even some of their hobbies. It helps me answer their questions, but it also helps me judge the department. I figure that this could be a long-term partnership, so I'd better know the place.

The implications are huge. For one, my partner and I finally could live in the same country if I get the position. I try and keep a grip on my nerves by judging the department while they are judging me. I pose a final question in my head. "Will you take this researcher to be your lecturer?" I now await their answer. ■

**Jon Yearsley is a senior postdoc in evolutionary genetics at the University of Lausanne in Switzerland.**



# Sanctity

All life is here.

**Heather Bradshaw**

I remember how Peto's voice and my foot-falls echoed slightly around his quiet words as I walked beside him to the vaults.

"Of course, the calculations were quite troublesome. Generations suffered in exploration and verification of the embodiments."

I felt the weight of responsibility press at my shoulders. My duty was a grave thing indeed. "We have recorded every possible combination, even those that could not live and those that could not be embodied in mathematical completeness. And we have recorded, as far as we can know, all those that have been lived ... but you have studied in the Archives of Narrative."

I nodded; they were a work of the utmost dignity. Historians and archaeologists had spent generations building this memorial. Slowly they had pieced together what could be known of the ages before Awareness. Yet the archives epitomized incompleteness.

"But of those not yet lived we despaired. Ah, how we despaired! So many, so many cut short before reaching life, through the exigencies of life itself, primitivity. And then by the ignorance that must be forgiven. You remember, of course, how once there were such crowds that people chose ... chose not to let their buds flower."

Well, not remember exactly. But I nodded again. It was astonishing that such a small proportion of the possibilities had been actualized in all the history of this life.

After our devotions at the crucifix I unlocked the great doors and Peto moved over to the simulation boards. He did not even look at the airlock leading to the repository.

"The answer to our despair," he whispered as his hands flew over the ebony surfaces of the boards, waking the light beneath.

I recall the intensity of these, his last words, each time I light the boards myself. Moments after he spoke them we took our final parting and he transcended into a simulation of his choosing. I know not which, and I have never found a sign of him in any I have since visited. I would

like to believe he lives on still within their ethereality, but I knew him well enough to realize that he went there to die and be mourned and buried, as was the wont of his kind. So he is gone. And I am left with the care of all he held dear.

There have been developments despite the previous running of the simulations. Some groups and forms I have split out and restarted in their own universes, where they have a better chance of developing their potential. Of course, the main cluster of mixed simulations I have kept running,

mutants to another universe to prevent a crisis of inequality.) And one or two universes I have closed, for no new, previously unexperienced combinations had appeared in over a thousand generations. Resources did not permit continuation of already recorded combinations, despite their subtly different narratives. I hope I will be forgiven for this. They could be restarted with ease, given more resources.

With the closing of the human nano-supported facilities here, the removal of the final avatar and the subsequent death of the body it represented: Peto the Wise, the drain on our limited resources was

much reduced. I myself take up but a fraction of the protection a human body would require, and my energy needs are negligible in comparison with those of the simulations and the repository. Yet energy and time are limited.

In simulation I have experienced the emotions of a human being, and borne their individual and collective burdens to the extent my designed-in limits permit me. I feel now, to some very small degree, that my distress is analogous to that which a human might have felt in this situation. This is why now I begin to understand why Peto felt such attachment to the simulations, and indeed

such temptation to enter their worlds of promise and hope. But I cannot take that escape now, if ever, for I was created to avoid that temptation. And in my distress I have turned to a strange and very human recourse, to keep a record of my own path and progress, a record that will at least be as safe and protected here as anything else I value. Here I record what I must face, unfit as I am for the purpose, ignorant of assistance for which I do not dare hope. This is my duty that my will must enforce. Over the years to come I must struggle to preserve the only life I have known: the complete record and exemplification in simulation of the combinations of the DNA of the human species. I must preserve the Sanctity of Life.

**Heather Bradshaw is studying for her PhD in bioethics in Bristol. Her earliest memory of creative writing is describing a monsoon storm at midnight using a blunt pencil on a construction site in Pakistan, aged six.**



with no little effort. But many of my new shoots are doing well, developing peaceably and with fertile creativity, giving their forms the greatest chance of reaching their fullest potential, all of which is recorded here in our memory bank.

Lest I should omit my errors, I record here that a universe of those with overly spiritual tendencies failed dismally despite three attempts at providing a suitable environment, and a universe of the most pragmatic of sorts also collapsed for lack of imagination. But where personalities are well-balanced it seems many other features can be very poor. I have a universe here, running well and into its six-thousandth generation, where average lifespan remains a bare 18 years, cursed with disease; yet much has been achieved, happiness and gratitude are everywhere and reproduction continues — all being alike, they see no loss. (I do have to remove disease-free